SOME NON-STANDARD APPROACHES TO THE STUDY OF SUMS OF HEAVY-TAILED DISTRIBUTIONS

By

Andrew William Richards

Submitted for the Degree of
Doctor of Philosophy
at Heriot-Watt University
on Completion of Research in the
Department of Actuarial Mathematics and Statistics
September 2009.

Contents

List of Figures

List of Tables

Acknowledgements

I am extremely grateful to Serguei Foss for always having faith in me. His help cannot be overestimated. I express my gratitude to Stan Zachary for many stimulating discussions and much sound advice. I would like to express my thanks to all my friends who rallied round, and particularly Simon Ellis and Marek Kwiatkowski, without whom I might never have had the courage to persevere.

Abstract

Heavy-tailed phenomena arise whenever events with very low probability have sufficiently large consequences that these events cannot be treated as negligible. These are sometimes described as low intensity, high impact events. Sums of heavy-tailed random variables play a major role in many areas of applied probability, for instance in risk theory, insurance mathematics, financial mathematics, queueing theory, telecommunications and computing, to name but a few areas. The theory of the asymptotic behaviour of a sum of independent heavy-tailed random variables is well-understood. We give a review of known results in this area, stressing the importance of some insensitivity properties of the class of long-tailed distributions. We introduce the new concept of the *Boundary Class* for a long-tailed distribution, and describe some of its properties and uses. We give examples of calculating the boundary class.

Geometric sums of random variables are a useful model in their own right, for instance in reliability theory, but are also useful because they model the maximum of a random walk, which is itself a model that occurs in many applications. When the summands are heavy-tailed and independent then the asymptotic behaviour has been known since the 1970s. The asymptotic expression for the geometric sum is often used as an approximation to the actual distribution, owing to the (usually) analytically intractable form of the exact distribution. However the accuracy of this asymptotic approximation can be very poor, as we demonstrate. Following and further developing work by Kalashnikov and Tsitsiashvili we construct an upper bound for the relative accuracy of this approximation. We then develop *new techniques* for the application of our analytical results, and apply these *in practice* to several examples. Source code

for the computer algorithms used in these calculations is given.

As we have said, the asymptotic behaviour of a sum of heavy-tailed random variables is well-understood when the random variables are independent, the main characteristic being *the principle of the single big jump*. However, the case when the random variables are dependent is much less clear. We study this case for both deterministic and random sums using a *novel approach*, by considering conditional independence structures on the random variables. We seek sufficient conditions for the results of the theory with independent random variables still to hold. We give several examples to show how to apply and check our conditions, and the examples demonstrate a variety of effects owing to the dependence, and are also interesting in their own right. All the results we develop on this topic are entirely new. Some of the examples also include results that are new and have not been obtainable through previously existing techniques. For some examples we study the asymptotic behaviour is known, and this allows us to contrast our approach with previous approaches.

Chapter 1

Introduction

Ever since the beginnings of statistical analysis statisticians have recognized that some data is 'intrinsically' bad. This is data that is not bad because of faults in the experimental design, or in the actual collection or recording of the data, but because of some property intrinsic to the data itself. It is bad because of the lack of an adequate model to describe the data, and hence the statistician does not have the requisite tools to analyze it. What constitutes 'bad data' depends on the interests of the analyst. The simplest situation is data that does not seem to be obeying the law of large numbers, and where empirical measures of the mean are not converging. A model describing this situation was first proposed by Simeon Poisson in his 1824 paper [70] in which he considers what is now known as the Cauchy distribution. Poisson demonstrated that the mean of the distribution does not exist, and that the (sample) mean of an independent sample of any size has exactly the same distribution as that of an individual observation; hence there is no concentration around the 'true' value of the mean. Poisson then goes on to dismiss the distribution as a monster, and one that has no practical purpose. Recent data on risk in insurance and financial institutions show that many risks, for instance operational risk, earthquake risks, nuclear power station risks, accident insurance are best modelled by infinite mean distributions [65]. Common interpretations of 'badness' are

- distributions with infinite or non-existent first moment, not obeying the law of

large numbers

- distributions with infinite second moment (often loosely described as infinite variance); these distributions do not (in general) obey the usual central limit theorem;

- distributions whose power moments are not all finite; these distributions have 'nonnormal' extreme behaviour;

- distributions that do not possess finite exponential moments; these distributions have nontypical tail behaviour for independent sums.

These types of distribution are variously described in the literature as *heavy-tailed, long-tailed, fat-tailed*, often with conflicting definitions (or no clear definition at all). Part of the difficulty lies in the different concept of 'badness' that different investigators have. In recent years a measure of consistency is beginning to arise in this area of terminology. This has been possible because the different categories listed above have a hierarchical structure with each successive group of distributions strictly including the previous ones.

The focus in this thesis is on sums of heavy-tailed random variables, and hence it makes sense to take the fourth most general category as our definition of the heavy-tailed distributions. All other definitions are then either equivalent or special cases of the definition given here. We shall use the term *long-tailed* to refer to a specific subclass of the heavy-tailed distributions; however, as all heavy-tailed distributions one is ever likely to meet are also long-tailed this should, in practice, cause no difficulties. We shall not use the term *fat-tailed* again. See section for precise definitions of these terms.

Later we shall give a more comprehensive list of commonly used heavy-tailed distributions. For now we observe that among these distributions are Pareto distributions, lognormal and Weibull with parameter less than 1.

1.1 Uses of heavy-tailed distributions

Heavy-tailed phenomena arise whenever events with very low probability have sufficiently large consequences that these events cannot be treated as negligible. These are sometimes described as low intensity, high impact events. The contrast with the normal distribution is instructive, where any event more than six standard deviations away from the mean is negligible.

Examples of these types of events abound in the insurance industry (e.g. claim sizes), finance industry (e.g. loss distributions) computing industry (e.g. file sizes on internet), engineering (e.g. power of wind currents), to name but a few. As an example, the following table shows the ten most costly insurance losses to the insurance industry as of 2006. These types of claims are very rare when compared to the vast number of ordinary claims, but for an insurer they may be greater than the combined sum of all other losses that year.

Date	Event	Insured losses* [US $ Billion]
2005, Aug 24	Hurricane Katrina	45
1992, Sep 23	Hurricane Andrew	22
2001, Sep 11	World Trade Center	21
2004, Jan 17	Northridge Earthquake	18
2004, Sep 2	Hurricane Ivan	12
2005, Sep 20	Hurricane Rita	10
2005, Oct 16	Hurricane Wilma	10
2004, Aug 11	Hurricane Charley	8
1991, Sep 27	Typhoon Mireille	8
1990, Jan 25	Storm Daria	7

Table 1.1: The ten costliest insurance events. *Excludes liability and life insurance. Source: SwissRe (2006)

Heavy tails arise in queueing systems (again, the example of file size distribution

is relevant), where they are often characterized by long-range dependence and self-similarity, which captures the phenomenon of burstiness at many different scales. Such models are not only of use in internet modelling, but also in finance, econometrics, hydrology, climate studies, linguistics, see, for instance [74] and references therein.

Heavy tails arise naturally in finance. If stock returns are supposed to follow a normal distribution, then stock prices will have a lognormal distribution. In fact, it has long been recognized, since the pioneering work of Mandelbrot on cotton prices [59] that assets can have much heavier tails than the lognormal. Mandelbrot proposed an α-stable model with infinite variance, and debate is still ongoing in the financial research community as to whether the second moment of asset prices is in fact infinite.

1.2 Classes of heavy-tailed distributions

The definitions given in this section will be sufficient to gain an overview of the work in this thesis, and to understand the motivation behind the definitions. As such we shall restrict ourselves in this section to considering non-negative random variables, with distribution functions F supported on the positive real axis $[0, \infty)$.

Definition 1.2.1. *A non negative random variable X has a heavy-tailed distribution function F if and only if*

$$\mathbf{E}(e^{tX}) = \int_0^\infty e^{tx} F(dx) = \infty \text{ for all } t > 0. \tag{1.1}$$

As we shall see in the next section, this definition implies that classical methods of dealing with problems such as the maximum of a random walk with negative drift by using an exponential change of measure are not applicable. However, in practice, (1.1) does not give enough information to make heavy-tailed problems analytically tractable. To this end we introduce the class of long-tailed distributions. Here, and throughout this work, we write the tail distribution function $\overline{F}(x) := 1 - F(x)$. Also, for distributions F, G with unbounded support (to the right) we say F and G are tail equivalent, and write $\overline{F}(x) \sim \overline{G}(x)$ for $\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{G}(x)} = 1$. For these and other notational definitions see Appendix A.

A slightly smaller class, but a subclass of the heavy-tailed distributions, is the class of subexponential distributions, and it is with this class that we shall be most interested during the course of this thesis. We write $F^{*2}(x) := \int_0^x \overline{F}(x-y)F(dy)$ for the two-fold convolution of F with itself, and, in general for $n \in \mathbb{N}$, $F^{*n}(x)$ for the n-fold convolution.

Definition 1.2.2. *A non negative random variable X has a subexponential distribution function F if and only if for any $n \in \mathbb{N}$*

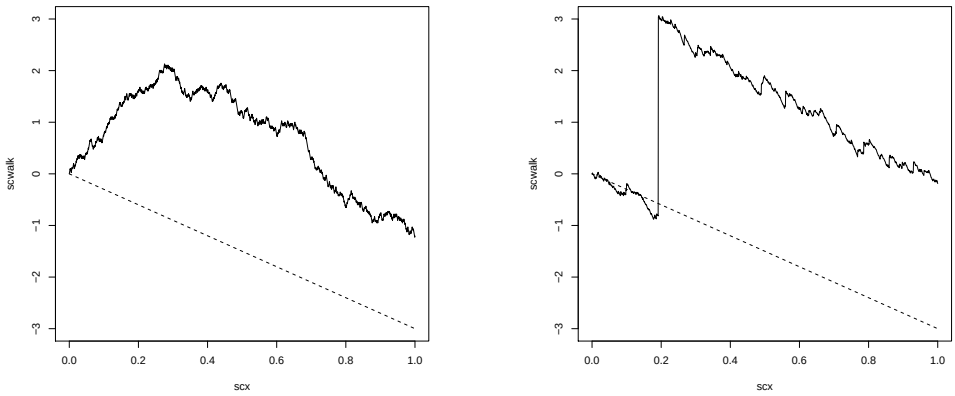
$$\overline{F^{*n}}(x) \sim n\overline{F}(x). \tag{1.2}$$

As we shall show in the next chapter, this definition is equivalent to the statement that, for two independent, identically distributed (i.i.d.) non-negative random variables (r.v.s) X_1, X_2 with common subexponential distribution function F, then, writing $x \vee y$ for $\max(x, y)$,

$$\mathbf{P}(X_1 + X_2 > x) \sim \mathbf{P}(X_1 \vee X_2 > x), \tag{1.3}$$

and similarly for a sum of n i.i.d. subexponential r.v.s.

This has the interpretation that for a sum of subexponential i.i.d. r.v.s to exceed a high level x, then asymptotically one of the r.v.s will exceed x, and the others will be 'small'. These statements will be made precise in the next chapter. As an illustration of this principle, which we refer to as the Principle of the Single Big Jump, in Figure 1.2 we show two random walks both scaled in time and space to give the same negative drift, and both conditioned to exceed the same high threshold. For the purposes of this illustration we relax the requirement that the random variables are non-negative, anticipating the definitions of Chapter 2. The first random walk has light-tailed increments (normally distributed) and the second has heavy-tailed (subexponential) increments (with Pareto distribution). In the light-tailed case the sample path follows to generally linear track to the threshold, then 'normal' drift takes over. In the subexponential case the principle of the single big jump is clear to see; it is also noteworthy that there is a significant overshoot in the subexponential case.



(a) Light-tailed (b) Heavy-tailed

Figure 1.1: Random walks conditioned to exceed high threshold

The subexponential class is a proper subclass of the long-tailed distributions, which is itself a proper subclass of the heavy-tailed distributions. We shall give examples of distribution functions belonging to one class but not another in the next chapter, but it will be seen that these distributions are essentially contrived. In practice, all the heavy-tailed distributions that one is likely to meet are subexponential. The table gives a list of commonly encountered heavy-tailed distributions, all of which are subexponential.

Name	Tail distribution or density	Parameters	Hazard rate
Pareto	$\overline{F}(x) = \left(\frac{\lambda}{\lambda+x}\right)^{\alpha}$	$\alpha, \lambda > 0$	$q(x) = \frac{\alpha}{\lambda+x}$
Burr	$\overline{F}(x) = \left(\frac{\lambda}{\lambda+x^{\gamma}}\right)^{\alpha}$	$\alpha, \lambda, \gamma > 0$	$q(x) = \frac{\alpha\gamma x^{\gamma-1}}{\lambda+x^{\gamma}}$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x-\mu)^2/(2\sigma^2)}$	$\mu \in \mathbb{R}, \sigma > 0$	$q(x) \sim \frac{\log x-\mu}{\sigma^2 x}$
Weibull	$\overline{F}(x) = e^{-cx^{\beta}}$	$c > 0, 0 < \beta < 1$	$q(x) = c\beta/x^{1-\beta}$

Table 1.2: Some common heavy-tailed distributions

Both the Pareto and Burr distributions belong to the family of regularly varying distributions which is the natural generalization of the Pareto distribution. For definitions of regular variation, and the related concepts of slow variation, rapid variation,

intermediate regular variation, etc., we refer the reader to the Appendix, or to the encyclopaedic reference [16].

1.3 An extended example

As an illustration of some of the issues we shall be concerned with, and to show the importance of the heavy-tail condition (1.1), we consider the Cramér model, introduced in [24], of ruin in insurance mathematics. Assume that an insurance company with initial capital u incurs claims $\{X_i\}_{i\geq 1}$ that arrive according to some renewal process $\{N(t)\}_{t\geq 0}$. We wish to find the probability that the company is ruined as a function of u. The size of the claims $X_1, X_2, \ldots$ are assumed to be i.i.d. non-negative random variables, and independent of the renewal process $\{N(t)\}$. Further, assume that income is generated as a deterministic continuous flow of premiums at rate c. Then the risk reserve at time $t \geq 0$ is given by

$$R(t) = u + ct - \sum_{n=1}^{N(t)} X_i, \tag{1.4}$$

where here, as throughout this thesis, we interpret the empty sum $\sum_{n=1}^{0} X_n = 0$. This is known as the Sparre-Andersen model. The probability of ruin, also known as the ruin function, is given by

$$\psi(u) = 1 - \mathbf{P}(R(t) \geq 0 \text{ for all } t \geq 0 | R(0) = u). \tag{1.5}$$

The claim surplus process is defined as $S(t) = u - R(t)$. Let $T_0 = 0, T_1, T_2, \ldots$ be the renewal epochs of the renewal process, with associated interarrival times $t_n = T_n - T_{n-1}$ for $n \geq 1$ and $t_0 = 0$. Note that $N(t) = \#\{k = 1, 2, \ldots : T_k \leq t\},\ t \geq 0$, where for a set A, $\#A$ is the cardinality of A. We consider the embedded random walk defined by sampling the claim surplus process at renewal epochs

$$S_n = \sum_{i=1}^{n} (X_i - ct_i),\ n = 0, 1, 2, \ldots, \tag{1.6}$$

where the losses between claims are $Y_i := X_i - c\tau_i$ are i.i.d. r.v.s with common distribution F. Let Y be an independent copy of Y_1. We assume $\mathbf{E}(Y) < 0$ (otherwise

the company will be ruined with probability 1). Hence we have a model of the surplus process as a random walk with negative drift. Let $M = \sup(0, S_1, S_2, \ldots)$ and $\tau(u) = \inf(n \geq 0 : S_n > u)$, with the convention that $\inf \emptyset = \infty$. Then the quantity of interest is

$$\psi(u) = \mathbf{P}(M > u) = \mathbf{P}(\tau < \infty). \tag{1.7}$$

Hence we are interested in the distribution of the maximum of a random walk with negative drift. This gives us the possibility of interpreting the ruin function as the tail distribution of the stationary waiting time of a first-come-first-served single server queue with general independent interarrival times and general independent service times (GI/GI/1), see, for instance [5].

To proceed with the classical analysis we now impose two further conditions (the Cramér conditions) on the distribution of Y over and above the net-profit condition $\mathbf{E}(Y) = -a$, where $a > 0$. First, let $M_Y(t) := \mathbf{E}(e^{tY})$ be the moment generating function for Y.

1. We assume there exists $\gamma > 0$ such that $M_Y(\gamma) = 1$.
   This condition implies that the distribution is not heavy tailed, and that there exists some $\gamma_u \in [\gamma, \infty]$ such that $\gamma_u = \sup(t : M_Y(t) < \infty)$.

2. We assume that the first derivative of M_Y satisfies $M'_Y(\gamma) < \infty$. (If $\gamma_u = \gamma$ we interpret the derivative as the left derivative.)

The existence of γ as in the first condition above allows us to introduce an exponential change of measure (also called an Esscher transformation) on the individual step size distribution F defined by

$$dF^*(x) = e^{\gamma x} dF(x), \tag{1.8}$$

which defines a new probability measure since

$$\int_{-\infty}^{\infty} dF^*(x) = \int_{-\infty}^{\infty} e^{\gamma x} dF(x) = M_Y(\gamma) = 1.$$

Denoting this transformed measure by $\mathbf{P}^*$, with corresponding expectation operator $\mathbf{E}^*$, we have

$$\mathbf{E}^*(Y) = \int_{-\infty}^{\infty} xe^{\gamma x} dF(x) = \mathbf{E}(Ye^{\gamma Y}) = M'_Y(\gamma),$$

which is positive and finite, and hence by simple renewal arguments we have, for any $u > 0$,

$$\mathbf{P}^*(\tau(u) < \infty) = 1. \tag{1.9}$$

Now let $A_n(u)$ be the set of sample path increments that first exceed level u on step n:

$$A_n(u) := \{(x_1, \ldots, x_n) : x_1 \leq u, \ldots, x_1 + x_2 + \cdots + x_{n-1} \leq u, x_1 + \cdots x_n > u\}.$$

We use $\mathbf{1}(A)$ to represent the indicator function of set A; that is

$$\mathbf{1}(x) = 1 \text{ if } x \in A; \quad \mathbf{1}(x) = 0 \text{ if } x \notin A.$$

Then

$$\begin{aligned}
\mathbf{P}(M > u) &= \mathbf{P}(\tau(u) < \infty) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(\tau(u) = n) \\
&= \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathbf{1}((x_1, \ldots, x_n) \in A_n(u)) dF(x_1) \ldots dF(x_n) \\
&= \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{-\gamma \sum_{i=1}^{n} x_i} \mathbf{1}((x_1, \ldots, x_n) \in A_n(u)) dF^*(x_1) \ldots dF^*(x_n) \\
&= \sum_{n=1}^{\infty} \mathbf{E}^*(e^{-\gamma S_n} \mathbf{1}(\tau(u) = n)) \\
&= \mathbf{E}^*(e^{-\gamma S_{\tau(u)}} \mathbf{1}(\tau(u) < \infty)) \\
&= \mathbf{E}^*(e^{-\gamma S_{\tau(u)}}),
\end{aligned}$$

where the last equality follows from (1.9).

Define the overshoot over level u as $B(u) = S_{\tau(u)} - u$, a quantity familiar from renewal theory as the forward recurrence time. It is well known, see, for example, [5], that if Y has a non-lattice distribution then the distributions of $B(u)$ converge weakly to a proper continuous distribution $B(\infty)$ as $u \to \infty$. In the case that Y has a lattice distribution with span h then $B(nh)$ converges to a proper discrete distribution as $n \to \infty$.

We then have

$$\mathbf{P}(M > u) = \mathbf{E}^*(e^{-\gamma B(u)})e^{-\gamma u}. \tag{1.10}$$

Since, for any value of $u > 0$ $B(u) \geq 0$ almost surely (a.s.), then we have Cramér's bound

$$\mathbf{P}(M > u) \leq e^{-\gamma u}, \tag{1.11}$$

and, if F is non-lattice, the following asymptotic result:

$$\mathbf{P}(M > u)e^{\gamma u} \to \mathbf{E}^*(e^{-\gamma B(\infty)}) \in (0, \infty) \text{ as } u \to \infty. \tag{1.12}$$

It is clear from this well-known analysis that a crucial role is played by the finiteness for positive values of the argument of the moment generating function (MGF) . This in turn implies that F is not heavy-tailed. The typical sample path of (most) non-heavy-tailed random walks which leads to exceeding a high level looks like a typical strong law of large numbers (SLLN) sample path under the appropriate probability measure. Indeed, one can find in, for instance, [38] that

$$\lim_{u\to\infty} \mathbf{P}(\sup_{n\leq\tau(u)} |S_n - an| < \varepsilon u | \tau(u), \infty) = 1$$

for any $\varepsilon > 0$

The intuitive reason that a similar argument will not work for heavy-tailed random variables is that the typical sample paths that exceed high levels do not resemble SLLN sample paths. We shall have more to say about typical sample paths in the next section.

The astute reader will have observed that there is a gap between the Cramér conditions and the case of heavy-tailed distributions, namely when the exponential moment is

finite for all $t < \gamma_u$, but either $M_Y(t) < 1$ for all $0 < t < \gamma_u$, or $M'(\gamma) = \infty$. These cases are considered in, for instance,[41, 54] and references therein.

The model described above depends critically on the independence assumptions about the X_i and $N(t)$. Often these assumptions are not realistic. To show how easily this independence may be lost, consider adapting the model to take into account a deterministic force of interest. Specifically, we consider the process which describes the risk reserve, with constant force of interest r, that is, after time t one unit of currency becomes e^{rt} units of currency. Then the risk reserve up to time t, $S_r(t)$ is given by

$$S_r(t) = ue^{rt} + \frac{c}{r}(e^{rt} - 1) - \sum_{i=1}^{N(t)} X_i e^{r(t-T_i)}.$$

Clearly the individual summands $X_i e^{r(t-T_i)} = X_i e^{r(t-(t_1+\cdots+t_i)}$ are not independent of each other, nor of the renewal process $N(t)$. it is not our intention in this thesis to consider this model in detail. Rather it is intended to provide motivation for studying sums of random variables in the heavy-tailed case, including those where dependence between the summands exists.

1.4 The Single Big Jump

For problems involving sums of independent random variables we want to know what is the most likely way that the sum becomes large asymptotically. For any positive independent random variables $X_1, X_2, \ldots, X_n$ with common distribution F,

$$\begin{aligned} \mathbf{P}(X_1 + \cdots + X_n > x) &\geq \mathbf{P}(\max(X_1, \ldots, X_n) > x) \\ &= 1 - F^n(x) \\ &\sim n\overline{F}(x), \end{aligned} \tag{1.13}$$

as $x \to \infty$. Hence, for any distribution with non-negative support, and for any $n \geq 1$,

$$\liminf_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \geq n, \tag{1.14}$$

where F^{*n} stands for the n-fold convolution of F with itself.

The main class of distribution functions that we shall be concerned with and that fail to meet the Cramér conditions will be the subexponential distributions. As we have already said, these are characterized by the fact that, asymptotically, a large value of an i.i.d. sum of subexponential random variables is exceeded when precisely one of the summands exceeds that level. If for the moment we focus on the case of two non-negative summands, X_1, X_2.

$$\begin{aligned}
&\mathbf{P}(X_1 + X_2 > x) = \mathbf{P}(X_1 \vee X_2 > x) + \mathbf{P}(X_1 \vee X_2 \leq x, X_1 + X_2 > x) \\
&= \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x) - \mathbf{P}(X_1 \wedge X_2 > x) + \mathbf{P}(X_1 \vee X_2 \leq x, X_1 + X_2 > x) \\
&:= \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x) - P_1(x) + P_2(x), \qquad (1.15)
\end{aligned}$$

where $X_1 \vee X_2 = \max(X_1, X_2)$ and $X_1 \wedge X_2 = \min(X_1, X_2)$. As we know from the definition of subexponentiality (1.2) that $\mathbf{P}(X_1 + X_2 > x) \sim 2\mathbf{P}(X_1 > x)$, and that because of the independence that $P_1(x)$ is negligible compared to $\mathbf{P}(X_1 > x)$, then it follows that $P_2(x)$ is negligible compared to $\mathbf{P}(X_1 > x)$. In that sense one summand is large and the other summand is, in some sense, small. In this case we say we have the *Principle of the single big jump*. If we relax the independence condition it may no longer be the case that $P_1(x)$ is negligible. If $P_2(x)$ is negligible, but $P_1(x)$ is not we shall say only that we have the *Principle of the big jump*.

1.5 Typical sample paths

1.5.1 A trichotomy

We shall be focusing on heavy-tailed behaviour. However, typically in the literature the distinction is drawn between heavy-tailed distributions, for which the exponential moment is infinite on the positive real axis, and light-tailed distributions, where $M_Y(t) < \infty$ for some $t > 0$. For the heavy-tailed distributions, many satisfy the condition for subexponentiality, and in fact this is true for all commonly encountered heavy-tailed distributions. In this case the typical sample path that leads to a high value of the sum exhibits the phenomenon of the single big jump. That is, one of

the (identically distributed, independent) summands is large and all the others are small. This is easily visualized in the case of two i.i.d. summands X_1, X_2 where, asymptotically, the contributions to the event $\{X_1 + X_2 > x\}$, for large x, come from $X_1 > x, X_2 \approx 0$ and $X_2 > x, X_1 \approx 0$, and everything else is negligible (the imprecise concept $X \approx 0$ will be made precise in the next chapter, although we shall not use this imprecise notation again). The intuition is that, as x increases, the bivariate distribution of (X_1, X_2) concentrates onto the two co-ordinate axes, and all other contributions are negligible.

In what is usually called the light-tailed case, the behaviour of the typical sample path is not so clear-cut. Again if we focus our attention on two (independent) summands, consider, for the sake of concreteness, two i.i.d. normal summands X_1, X_2. It is well known that the typical sample path that leads to the extreme event $\{X_1 + X_2 > x\}$, for large x, comes from $\{X_1 \approx \frac{1}{2}x, X_2 \approx \frac{1}{2}x\}$, so the contributions concentrate around the diagonal $\{X_1 = X_2, X_1 + X_2 > x\}$. However, if we consider two i.i.d. exponential summands, there is not this concentration effect, and all possible types of path contribute non-negligibly to the sum.

This suggests that the non-heavy-tailed case should be split into those cases that have 'super-exponential' behaviour, where the contributions to the i.i.d. sum are concentrated on the diagonal, and those distributions whose tails are exponential-like and where asymptotically the contributions are more spread out. It is the super-exponential case that we suggest should be called light-tailed. However, to avoid confusion, we shall use the (inelegant) term non-heavy-tailed to describe distributions for which the mgf is finite for some positive value of the argument.

There are many open questions about the behaviour of i.i.d. sums in the non-heavy-tailed cases, particularly in the exponential-like class. A variety of behaviours is possible in this class apart from the true exponential behaviour, and the class includes the classes $\mathcal{L}(\gamma)$ and $\mathcal{S}(\gamma)$ introduced in [19]. We will not have time to discuss these in this thesis. We shall, however, use the idea of the concentration of the multi-variate distributions in a heuristic way later to give guidance to our thinking about dependent

sums.

1.5.2 Big Jumps and the CLT

It is commonly believed in some areas of applications that large sums of finite variance i.i.d. random variables are well-approximated by normal distributions through the use of the classical central limit theorem. This leads to practice, typically found in some areas of engineering and economics, for example, that when a phenomenon has many small additive causes, that the individual nature of those causes need not be investigated too closely, as the normal approximation can be used instead. Clearly, in the case of finite variance, the central limit theorem holds; however, it is possible to misunderstand the nature of the weak convergence, and to expect the normal approximation to do too much work. Large deviations sample paths of heavy-tailed (subexponential) i.i.d. random variables do not resemble those of normal sums, regardless of how many summands there are, and the phenomenon of the single big jump will always lead to an underestimate of the probability of a large deviation if the normal approximation is used. Essentially this is because in the heavy-tailed case one is not entitled to change the order of the limiting processes in the expression $\lim_{x\to\infty}\lim_{n\to\infty}\mathbf{P}(S_n/n > x)$. The normal approximation works well in any finite region as the number of summands increases, but it never captures the tail behaviour correctly.

For exactly the same sort of reasons one should be careful when using (general) phase-type distributions to approximate heavy-tailed distributions. Phase-type distributions are dense (in the sense of weak convergence) in the space of probability distributions on $(0,\infty)$, see for example [5] Chapter III, Theorem 4.2, but as they are (Markov modulated) mixtures of exponential distributions they inherit the tail behaviour of exponential-like distributions. Weak convergence ensures that, as the number of phases increases, these phase type distributions approximate any non-negative distribution well in any finite region, but the heavy-tail behaviour is never captured.

1.6 Rare Event Simulation

We shall be interested in the tail of the distribution of a sum of random variables when some or all of the random variables are heavy-tailed, and in particular in the probability of the sum exceeding some high level. If we take the simplest situation in which the random variables are independent and identically distributed, it nonetheless remains the case that the distribution of the sum, calculated by the convolution of the individual distribution functions, is analytically intractable. This makes Monte Carlo simulation seem an attractive prospect.

Let $A(x)$ be some event depending on a parameter x in such a way that $\mathbf{P}(A(x))$ decreases monotonically to 0 as $x \to \infty$. Crude Monte Carlo estimation of $A(x)$ consists of simulating i.i.d. replicates $Z_1(x), Z_2(x), \ldots, Z_N(x)$ of the random variable $Z(x) = \mathbf{1}(A(x))$ and estimating the probability $\mathbf{P}(A(x))$ by the empirical mean $\widehat{M}(x) = (Z_1(x) + \cdots + Z_N(x))/N$. This is an unbiased estimator with variance $\mathbf{V}(\widehat{M}(x)) = \mathbf{P}(A(x))(1 - \mathbf{P}(A(x)))/N)$, which tends to 0 as the event A becomes rarer. However, when the event $A(x)$ is a rare event it is not sufficient that the variance is small; we need the relative precision of the estimate

$$\frac{\sqrt{\mathbf{V}(\widehat{M}(x))}}{\mathbf{P}(A(x))} = \sqrt{\frac{1 - \mathbf{P}(A(x))}{N\mathbf{P}(A(x))}}$$

to be small as well. This relation demonstrates clearly that, in order to achieve a given level of relative precision, then the sample size in crude Monte Carlo must increase (asymptotically) linearly with the inverse of the probability of the event $A(x)$ as $A(x)$ becomes rarer.

Hence we seek adaptations of the Monte Carlo method which will reduce the variance of the estimate. A commonly used criterion for the efficiency of a simulation method is that of *logarithmic efficiency*. Ideally one would like the variance of each Monte Carlo estimator $\widehat{M}(x)$ to decrease to 0 fast enough. In particular one would wish that for all x

$$\frac{\mathbf{V}(\widehat{M}(x))}{(\mathbf{P}(A(x)))^2} < \infty.$$

This condition is known as *bounded relative error* and it implies that the number of replications required to obtain a given level of relative accuracy is bounded in x. However the condition places restrictions on the choice of method that are often unrealistic in practice. Logarithmic efficiency is the compromise, where one requires that $\mathbf{V}(\widehat{M}(x))$ goes to zero at least as fast as $\mathbf{P}(A(x))$ raised to a power arbitrarily close to the second power:

$$\limsup_{x\to\infty} \frac{\mathbf{V}(\widehat{M}(x))}{(\mathbf{P}(A(x)))^2} < \infty.$$

This then implies that the number of replications grows at a rate $o(-\log \mathbf{P}(A(x))$.

Several *variance reduction* techniques exist to improve efficiency of Monte Carlo simulation for rare events, the most important of which is importance sampling (see, for instance [8] and references therein). In importance sampling one simulates, not from the natural probability measure $\mathbf{P}$, but from a different probability measure $\tilde{\mathbf{P}}$ which has the property that there is a random variable $L(x)$ such that the random variable $Z(x) = L(x)\mathbf{1}(A(x))$ satisfies

$$\mathbf{P}(A(x)) = \tilde{\mathbf{P}}(Z(x)).$$

Specifically let us consider the event $A(x) = \{M > x\}$ of the supremum of a sum of i.i.d. random variables with negative mean value and common distribution F exceeding level x. If F satisfies the Cramér conditions (which in particular imply that F is not heavy-tailed), then the analysis of Section 1.3 suggests that we use the exponentially twisted change of measure defined in (1.8). In other words, we are choosing a change of measure so that the rare event occurs with probability 1 under the new measure, and the law of large number behaviour under $\tilde{\mathbf{P}}$ copies the large deviation behaviour under $\mathbf{P}$. We can take $L(x)\mathbf{1}(A(x))$ to be $e^{-\gamma B(x)}e^{-\gamma x}$ as in (1.10) (since under the new measure $A(x) > 0$ with probability 1). Then it is known that this exponential change of measure gives an importance sampling scheme that is logarithmically efficient (see [8] and references therein).

However, the method of exponential twisting is not available in the heavy-tailed case. As discussed in the previous section this is not surprising as one cannot find a change of

measure where the law of large numbers behaviour mimics the large deviations single big jump behaviour under the natural measure. The exponential change of measure importance sampling scheme is a state-independent scheme because the individual random variables $Z_1, Z_2, \ldots$ are all sampled under the same changed measure. In [14] it is shown that no state-independent efficient schemes exist when the distribution F is regularly-varying.

Various schemes have been proposed to try to deal with this problem. The first logarithmically efficient simulation algorithm was developed in [6] and was based on a conditional Monte Carlo approach, conditioning on the order statistics of the replications $Z_1, Z_2, \ldots$. However this scheme was shown to be logarithmically efficient for F regularly varying but not for Weibull-type tails. Subsequently in [7] a scheme was proposed that is logarithmically efficient for both regularly varying and Weibull-type tails. This new scheme involved perhaps the most important class of methods: the state-dependent importance sampling methods. The individual replications are sampled under changes of measure where the change of measure selected depends on the previous history of the sampling procedure. This allows the scheme to mimic the single big jump behaviour, making it more likely that a large increment will occur next if no such large increment has occurred before. In [48] a logarithmically efficient scheme was proposed, based on state-dependent twisting of the hazard rate. More recently Blanchet and Glynn [17] have introduced a scheme that has bounded relative error for both light and heavy-tails.

1.7 Overview of the thesis

In Chapter 2 we give an introduction to heavy-tailed distributions. The majority of the results in this chapter are not new. We focus on the asymptotics of the tail of the distribution of the sum of n independent random variables under the assumptions of heavy-tailedness, long-tailedness and subexponentiality. For long-tailed distributions we review the concept of h-insensitivity. We review necessary and also sufficient

conditions for subexponentiality, and consider closure properties of the class of subexponential distributions. We also review a useful technical lemma, Kesten's lemma, and introduce a new result, Lemma 2.2.1 that allows us to control the asymptotic behaviour of two quantities that, in some sense, measure how long and how subexponential a distribution is. We then introduce the concept of the Boundary class for a long-tailed distribution; the results of this section are all new and are based on results in [39]. Although not all long-tailed distributions possess a Boundary class, nonetheless all commonly used distributions that are heavy-tailed (but not slowly varying) do possess one, and, in the case of distributions with well-defined hazard rates, the boundary class is easy to construct, and is generated (in some sense) by any one of its members. The importance of the boundary class lies in the fact that many results rely on the existence of a function for which the distribution is h-insensitive and which satisfies some condition. Such a function may be difficult to find: and the fact that a particular function for which the distribution is h-insensitive but which does not satisfy the condition does not mean that a different function may not suffice. We show that if the property is satisfied by all multiples of any generator of the boundary class then there exists some function for which the distribution is h-insensitive and which satisfies the condition.

In Chapter 3 we consider the problem of finding an upper bound for the geometric sum of independent heavy-tailed distributions. This chapter is based on [72]. Geometric sums are important in many fields in their own right, particularly in reliability problems, but also model the maximum of a negative drift random walk, which in turn is useful in insurance, credit risk management and queueing problems, to mention a few among many areas. There is a standard asymptotic expression for the tail of the geometric sum which is often used as an approximation. We show how inaccurate this asymptotic approximation can be. The analytical results in this chapter are based on an idea in [50, 51] which we have reformulated and extended. In these papers Kalashnikov and Tsitsiashvili developed an upper bound for the relative error, but their result was not entirely correct. We use the same general approach but our results are

expressed in a more probabilistic light, and this allows us to include the term that was omitted in [51], and also to tighten the bound. The method is based first on an estimation technique based on choosing a suitable test function; the test function is then multiplied by a suitably calculated constant to provide the bound. The calculation of the constant involves some numerical calculation of the actual relative error over some finite region. Kalashnikov and Tsitsiashvili found that in many cases the finite region over which the actual relative error had to be calculated was so large as to be practically unfeasible. The same applies to many cases in our analysis. However, we were able to find alternative methods of constructing the test function which allowed the construction of an upper bound in all cases that we tried. We give examples of applying the method to geometric sums where the individual summands have Pareto or Weibull tails. We also give code for performing the necessary calculations, which we present in an appendix. We note that the algorithms have been improved since the work in [72], and hence the numerical results have been slightly tightened.

In Chapter 4, which is based on [39], we consider the problem of finding the asymptotics for the tail of the distribution of the sum of heavy-tailed random variables. All results in this chapter are new, although some of the examples have been studied before. We do not assume the random variables are independent, not that they are identical. The aim is to find conditions on the strength of the dependence that guarantees that the standard results from the theory of independent heavy-tailed random variables carry over to the dependent case. Our approach, which has not been studied in the context of heavy-tailed dependence before, is to describe the dependence structure by means of a conditional sigma algebra, such that conditional on that sigma algebra the random variables are independent. We first consider conditions for sums of non-negative dependent random variables. The main results in this chapter are the formulation of Conditions (1-6) and Theorems 4.4.1 and 4.4.2 and Lemma 4.4.1. We give some instructive examples demonstrating how to apply these conditions. Then we discuss the case of random variables with distributions supported on the whole real line. This involves the development of an extra condition (7), and we state and

prove Theorem 4.6.1. Finally we consider some heuristic methods which in some cases allow us to derive results without recourse to the whole of our machinery.

Chapter 2

Heavy-tailed distributions

The first chapter of this thesis served as a general introduction to motivate the topics to come. This chapter is concerned with the mathematical details of heavy-tails that will be of use to us in the study of these problems. Most of the results are well-known and indeed some date back to the introduction of the concepts of subexponentiality and long-tailedness by Chistyakov in 1964 [22] (although he did not use those particular terms). However, some of the results in this chapter are new, namely Lemma 2.2.1 and also all the concepts and results in Section 2.3 on the Boundary Class. Many of the well-known results have previously been proved in a mostly analytical style. Proofs that emphasize the probabilistic nature of the argument help strengthen intuition, and as such many of the propositions have been taken from the monograph [37] which is itself a survey, and contains references to the original papers. We begin by giving general definitions of heavy-tailed, long-tailed and subexponential distributions. We then introduce the well-known device of h-insensitivity which enables us to prove many results in a more probabilistic light. We discuss closure properties of subexponential random variables under addition (note that we shall interchangeably say that a random variable is subexponential, or heavy- or long-tailed, when the distribution of the random variable is). Finally we consider some known results on subexponentiality of products of random variables, and consider how we can gain insight into some of these results in the light of the previous section.

2.1 Heavy and Long Tails

In this section we introduce some basic definitions and results concerning heavy-tailed and long-tailed distributions. We give our definitions in terms of real valued random variables with unbounded support, and focus on the right tail of the distribution. Definitions for the left tails can easily be constructed. Where there is no danger of confusion (that is almost everywhere throughout this thesis) we shall take the term heavy-tailed to mean right heavy-tailed, and so forth for light-tailed. We note that the properties of heavy- or long-tailedness are indeed tail properties of the distribution in the sense that, for any x, they depend only on the restriction of the distribution to (x, ∞).

Let $X, X_i,\ i \geq 1$, be independent random variables, with unbounded support, and with common distribution function $F(x) := F(-\infty, x] = \mathbf{P}(X \leq x)$, and tail distribution function $\overline{F}(x) := F(x, \infty) = \mathbf{P}(X > x)$. Then

Definition 2.1.1. *A distribution function F (or the random variable X) has a (right) heavy-tail if, for all $t > 0$,*

$$\mathbf{E}(e^{tX}) = \infty. \tag{2.1}$$

It is known that (see, for example [37] and references therein)

Proposition 2.1.1. *The distribution function F is heavy-tailed if and only if* $\limsup_{x\to\infty} e^{tx}\overline{F}(x) = \infty$ *for all $t > 0$.*

The lim sup (from now on we shall assume all limiting relations are as $x \to \infty$, unless it is stated to the contrary) in the previous proposition suggests that the class of heavy-tailed distributions, although most closely connected to the failure of the Cramér conditions, is not in practice the most practical class of distributions to work with. A more natural condition for 'real-life' distributions to obey would be $\lim e^{tx}\overline{F}(x) = \infty$ for all $t > 0$.

In terms of the single big jump idea, for the sum of two i.i.d. heavy tailed random variables all we can say is given by the following proposition, see, for example, [37] and references therein.

Proposition 2.1.2. *For any heavy-tailed distribution F on $\mathbb{R}^+$,*

$$\liminf_{x\to\infty} \frac{\overline{F*F}(x)}{\overline{F(x)}} = 2. \tag{2.2}$$

Not only is the restriction of this proposition to those distributions supported on the positive real line unwanted, but neither is the converse of this proposition true. In [36] we find an example of a distribution on $\mathbb{R}^+$ that is not heavy-tailed yet for which

$$\liminf_{x\to\infty} \frac{\overline{F*F}(x)}{\overline{F(x)}} = 2.$$

To begin to address this problem we introduce the class of long-tailed (to the right) distributions.

Definition 2.1.2. *A distribution function F (or the random variable X) has a (right) long-tail if, for all $t \in \mathbb{R}$*

$$\lim_{x\to\infty} \frac{\overline{F}(x+t)}{\overline{F}(x)} = 1. \tag{2.3}$$

If we reinterpret the definition probabilistically in terms of a random variable X with distribution F we see that it is equivalent to

$$\mathbf{P}(X > x+t | X > x) \to 1 \text{ as } x \to \infty, \text{ for } t > 0.$$

This has the colloquial interpretation that if a situation is bad, it is almost certainly worse.

Also we observe that we can rewrite the definition for $t > 0$ as

$$\lim_{x\to\infty} \frac{F(x, x+t]}{F(x,\infty)} = 0, \tag{2.4}$$

and a similar expression for $t < 0$.

Before continuing we remind the reader that a positive function l is slowly varying (at infinity) if, for all $\lambda > 0$, $l(\lambda x) \sim l(x)$.

Proposition 2.1.3. *A distribution F is long-tailed if and only if the function $g(x) = \overline{F}(\log x)$ is slowly varying (at infinity).*

Proof. From the definition of g we have, for any $\lambda > 0$,

$$g(\lambda x) = \overline{F}(\log x + \log \lambda) \sim \overline{F}(\log x) = g(x).$$

□

It is well known (see, for example, [37] and references therein) that the class of long-tailed distributions is a proper subclass of the heavy-tailed distributions, and indeed it is easy to show that

Proposition 2.1.4. *For any long-tailed distribution F and any $t > 0$,*

$$\lim_{x\to\infty} e^{tx}\overline{F}(x) = \infty.$$

Equation (2.4) gives us the indication of how to construct a heavy-tailed distribution that is not long-tailed. We simply need to construct a distribution F that stochastically dominates a heavy-tailed distribution, yet infinitely often, at $x_1, x_2, \ldots$, has 'concentrations' (that is, not necessarily discrete jumps) of probability mass that are not negligible compared to $\overline{F}(x_1), \overline{F}(x_2), \ldots$. As an example consider distribution functions $\overline{F}_1(x) = \min(1, x^{-\alpha})$ and $\overline{F}_2(x) = \min(1, 2x^{-\alpha})$ for $\alpha > 0$. Define a sequence of points by $x_1 = \sup\{x : \overline{F}_2(x) = 1\}$, and, for $n > 1$, $\overline{F}_2(x_n) = \overline{F}_1(x_{n-1})$. Then, for $x \geq x_1$, we define our distribution function F by

$$\overline{F}(x) = \overline{F}_1(x_n) \text{ for } x_n \leq x < x_{n+1}.$$

For completeness sake we can proscribe that $\overline{F}(x) = 1$ for $x < x_1$, but long- and heavy-tailedness are tail properties, so from that point of view it is not really necessary.
In terms of convolutions we have now strengthened (1.14) (again, see, for example, [37] and references therein):

Proposition 2.1.5. *Let the distribution F be long-tailed. Then for any $n \geq 1$*

$$\liminf_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \geq n. \tag{2.5}$$

However, to obtain the regularity under convolution that we desire we have to strengthen the definitions we have already encountered to obtain a further proper subclass of the long-tailed distributions, the subexponential distributions, first introduced by Chistyakov [22].

First, however, we introduce another idea that will be very useful in future analysis, particularly when considering sums of random variables and convolutions of distributions. This is the concept of h-insensitivity. The idea can be traced back to Borovkov [18], but the first complete development of this idea that we can find in the context of long-tailed distributions is in [37].

Proposition 2.1.6. *A distribution F is long-tailed if and only if there exists a strictly positive non-decreasing function h such that $h(x) \to \infty$ and*

$$\overline{F}(x + h(x)) \sim \overline{F}(x) \quad \text{as } x \to \infty. \tag{2.6}$$

We then say that the long-tailed distribution F is h-insensitive.

Proof. Clearly if h is a strictly positive non-decreasing function tending to ∞, and F is such that (2.6) is satisfied, then F is long-tailed.

For the other implication, assume F is long-tailed. Take $x_0 = 1$. For any integer $n \geq 1$ we can choose x_n such that $x_n - x_{n-1} \geq 1$ and

$$F(x, x + n] \leq \overline{F}(x)/n \text{ for all } x > x_n. \tag{2.7}$$

For $x \in (x_n, x_{n+1}]$ define $h(x) = n$. For $x \leq x_0$ we can define $h(x) = 1$. By construction $h(x) \to \infty$ as $x \to \infty$ and $\overline{F}(x + h(x)) \sim \overline{F}(x)$. □

Note that be simple adaptations of (2.7) and adapting the step-function h to be a piecewise linear function we can choose h to be continuous or continuous and concave if so desired. Also note that it is only the tail of the function h that is going to be of importance and so we may only wish to require that the function is, for instance, eventually non-decreasing or eventually concave.

Observe that if a distribution F is h-insensitive for some positive non-decreasing function h_1 tending to ∞, and h_2 is another such function for which $h_2(x) \leq h_1(x)$ eventually, then F is also h_2-insensitive.
As a final observation notice that we can use the 'negative' version of (2.3) to obtain

Proposition 2.1.7. *A distribution F is long-tailed if and only if there exists a strictly positive non-decreasing function h such that $h(x) \to \infty$ and*

$$\overline{F}(x - h(x)) \sim \overline{F}(x).$$

2.2 Subexponential Distributions

In practical usage we expect our distributions to exhibit reasonably good tail behaviour. All heavy-tailed distributions in common use, from Pareto distributions through their relatives in the regularly varying and intermediately varying distributions (see the appendix for definitions of these), to lognormal distributions and heavy Weibull distributions have well-behaved tail behaviour in the sense that i.i.d. sums follow the principle of the single big jump, and large values of the sum are dominated by one, and precisely one, of the summands. We shall give a general definition for subexponentiality of distributions supported on the whole real line, but in practice much of our focus will be on distributions on the positive real axis. We begin by considering distributions on $\mathbb{R}^+$, and by rephrasing the definition from (1.2)

Definition 2.2.1. *A distribution F with unbounded support on the positive real axis is subexponential if*

$$\overline{F^{*2}}(x) \sim 2\overline{F}(x).$$

The definition in (1.2) then follows straightforwardly by induction. Probabilistically, in terms of non-negative i.i.d. random variables X_1, X_2 with common subexponential

distribution F we have

$$\begin{aligned}\mathbf{P}(X_1 + X_2 > x) \sim & \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x) \\ = & \mathbf{P}(\max(X_1, X_2) > x) - \mathbf{P}(X_1 > x)\mathbf{P}(X_2 > x) \\ \sim & \mathbf{P}(\max(X_1, X_2) > x).\end{aligned}$$

This demonstrates the principle of the single big jump.

To understand the relationship between the classes of long-tailed and subexponential distributions we use the decomposition

$$\begin{aligned}\mathbf{P}(X_1 + X_2 > x) = & \mathbf{P}(X_1 > x) + \mathbf{P}(X_1 + X_2 > x, X_1 \leq x) \\ = & \mathbf{P}(X_1 > x) + \int_0^x \overline{F}(x - y) dF(y). \end{aligned} \tag{2.8}$$

Before proceeding we note that we shall use Landau's little-o notation. Let $a(x)$ and $b(x)$ be two positive functions such that

$$0 \leq l_1 = \liminf_{x \to \infty} \frac{a(x)}{b(x)} \leq \limsup_{x \to \infty} \frac{a(x)}{b(x)} = l_2 \leq \infty.$$

We write $a(x) = O(b(x))$ if $l_2 < \infty$ and $a(x) = o(b(x))$ if $l_2 = 0$. We say that $a(x)$ and $b(x)$ are weakly tail equivalent, written $a(x) \asymp b(x)$, if both $l_1 > 0$ and $l_2 < \infty$.

Proposition 2.2.1. *Let F be a subexponential distribution on $\mathbb{R}^+$. Then F is long-tailed.*

Proof. From (2.8) and the subexponentiality of F we have

$$\int_0^x \overline{F}(x - y) dF(y) - \overline{F}(x)F(x) = o(\overline{F}(x)) \text{ as } x \to \infty.$$

Take any $t > 0$. Then for any $x > t$

$$\int_0^x \overline{F}(x - y) dF(y) - \overline{F}(x)F(x) \geq F(x - t, x]F(t, x].$$

Hence $F(x - t, x] = o(\overline{F}(x))$, which is equivalent to the long-tailedness of F. □

Hence the subexponential class is a subclass of the long-tailed class. It is also a proper subclass. We shall not provide a specific example, but note that if we rewrite (2.8) and take the limit as $x \to \infty$, we get

$$\lim_{x\to\infty} \frac{\mathbf{P}(X_1 + X_2 > 0)}{\mathbf{P}(X_1 > x)} = 1 + \lim_{x\to\infty} \int_0^x \frac{\overline{F}(x-y)}{\overline{F}(x)} dF(y),$$

and then subexponentiality follows from long-tailedness if it is possible to justify the interchange of limit and integration. The fact that it requires some ingenuity to construct an example where the interchange is not valid gives strong evidence to the idea that it is the subexponential distributions that are, in some sense, the natural heavy-tailed distributions.

To extend the definition of subexponentiality to general distributions defined on $\mathbb{R}$ we need to add some extra condition. For an example of a distribution which is not heavy-tailed yet satisfies (1.2) see, for example, Example 3.3. in [37].

Definition 2.2.2. *Let F be a distribution on $\mathbb{R}$ such that $\overline{F}(x) > 0$ for all x. Then F is subexponential if F is long-tailed and*

$$\overline{F^{*2}}(x) \sim 2\overline{F}(x).$$

There is an alternative way to define subexponentiality on the whole line. If we consider a random variable X with distribution F, then we define (here and throughout) $X^+ = \max(0, X)$ with distribution F^+, so that $F^+(x) = 0$ for $x < 0$ and $F^+(x) = F(x)$ otherwise. Then we can say that F is subexponential on the whole line if F^+ is a subexponential distribution on $\mathbb{R}^+$. These two definitions are equivalent as shown in the next proposition.

Proposition 2.2.2. *Let F be a distribution on $\mathbb{R}$. Then the following are equivalent:*

i) F is long-tailed and $\overline{F^{*2}}(x) \sim 2\overline{F}(x)$;

ii) the distribution F^+ is subexponential.

Proof. Let X_1, X_2 be independent random variables with common distribution F. First assume F is long-tailed and (1.2) holds. Let h be a positive non-decreasing function tending to ∞ such that F is h-insensitive. Then for $x > 0$

$$\begin{aligned}\mathbf{P}(X_1^+ + X_2^+ > x) \leq & \mathbf{P}(X_1 + X_2 > x - h(x)) + \mathbf{P}(X_1 > x, X_2 < -h(x)) \\ & + \mathbf{P}(X_2 > x, X_1 < -h(x)) \\ \leq & \overline{F}(x - h(x)) + 2\overline{F}(x)F(-h(x)).\end{aligned}$$

So, by long-tailedness of F and the fact that $F^+(x) = F(x)$ for $x \geq 0$,

$$\limsup_{x\to\infty} \frac{\overline{F^+ * F^+}(x)}{\overline{F^+}(x)} \leq 2.$$

This, together with (1.14), implies that F^+ is subexponential.
For the converse assume that F^+ is subexponential. Hence by Proposition 2.2.1 F is long-tailed, and so is F^+. Also $X_1 + X_2 \leq X_1^+ + X_2^+$, so

$$\overline{F * F}(x) \leq \overline{F^+ * F^+}(x) \sim 2\overline{F^+}(x)$$

as $x \to \infty$. So, by (2.5), $\overline{F * F}(x) \sim 2\overline{F}(x)$ as required. □

The following theorem is key to understanding how the principle of the single big jump implies that sums where no jumps are 'small' are asymptotically negligible.

Theorem 2.2.1. *Let F be a long-tailed distribution on $\mathbb{R}$, and let X_1, X_2 be independent random variables with common distribution F. Then the following are equivalent:*

i) F is subexponential;

ii) there exists a positive non-decreasing function h tending to ∞ such that $h(x) < x/2$ and such that F is h-insensitive and

$$\mathbf{P}(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)) = o(\overline{F}(x)) \quad \text{as } x \to \infty; \tag{2.9}$$

iii) for every positive non-decreasing function h tending to ∞ with $h(x) < x/2$ relation (2.9) holds.

Proof. i) $\Leftrightarrow$ ii). Let h be as described in ii). We use the decomposition

$$\begin{aligned}\mathbf{P}(X_1 + X_2 > x) =& \mathbf{P}(X_1 + X_2 > x, X_1 \leq h(x)) + \mathbf{P}(X_1 + X_2 > x, X_2 \leq h(x)) \\ &+ \mathbf{P}(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)). \end{aligned} \tag{2.10}$$

Now,

$$\mathbf{P}(X_1 + X_2 > x, X_1 \leq h(x)) \leq \mathbf{P}(X_2 > x - h(x)) \sim \overline{F}(x),$$

and

$$\mathbf{P}(X_1 + X_2 > x, X_1 \leq h(x)) \geq \mathbf{P}(-h(x) < X_1 \leq h(x), X_2 > x + h(x)) \sim \overline{F}(x).$$

and similarly for the second term on the right of (2.10). Hence the condition of subexponentiality is equivalent to (2.9).

ii)$\Leftrightarrow$ iii). Clearly iii) $\Rightarrow$ ii). The converse follows by consideration of the remarks following Proposition 2.1.6. $\square$

Relations (2.4) and (2.9) and the rate at which they converge give us an indication of, respectively, how long-tailed and how subexponential a distribution is. For any distribution F we define:

$$K_F(x, r) := \frac{F(x - r, x]}{\overline{F}(x)}, \tag{2.11}$$

$$J_F(x, r) := \int_r^{x-r} \frac{\overline{F}(x - y)}{\overline{F}(x)} F(dy). \tag{2.12}$$

These notations are related to similar quantities introduced in [51].

For a long-tailed distribution F

i) $K_F(x, r) \to 0$ as $x \to \infty$ for fixed r;

ii) $K_F(x, r)$ is monotonically increasing in r for fixed x;

while for a subexponential distribution F

iii) $J_F(x, r) \to \overline{F}(r)$ as $x \to \infty$ for fixed r;

iv) $J_F(x, r)$ is monotonically decreasing in r for fixed x.

If we let F be a subexponential function and h be a positive non-decreasing function tending to ∞ such that F is h-insensitive, then we shall also define:

$$K_{F,h}(x) := \frac{F(x-h(x),x]}{\overline{F}(x)}, \tag{2.13}$$

$$J_{F,h}(x) := \int_{h(x)}^{x-h(x)} \frac{\overline{F}(x-y)}{\overline{F}(x)} F(dy). \tag{2.14}$$

Then, by (2.4) we have

$$K_{F,h}(x) = o(1), \tag{2.15}$$

and by (2.9)

$$J_{F,h}(x) = o(1). \tag{2.16}$$

However, we can go further than this. For a given subexponential distribution F we can choose a suitable function h so that F is h-insensitive and we can also control the rate of convergence of $J_{F,h}(x)$ and $K_{F,h}(x)$ to zero. The following lemma is new, but follows an idea in [51].

Lemma 2.2.1. *Let F be a subexponential distribution on the whole line. For any $c > 1$ there exists a function $h(x)$ (the choice of which depends on c) and a constant $x_1 \equiv x_1(c) > 0$ such that, for $x \geq x_1$,*

i) $h(x)$ is an (eventually) positive non-decreasing concave function tending to ∞;

ii) $J_{F,h}(x) \leq c\overline{F}(h(x))$;

iii) $K_{F,h}(x) \leq \overline{F}(h(x))$.

Proof. This proof uses the same approach as that of Proposition 2.1.6.
Let F be subexponential on the whole line and $c > 1$. Then there exists $k \in \mathbb{N}$ such that

$$k = \min\left(j \in \mathbb{N} : \inf_{x \geq j} \frac{\overline{F}(x+1)}{\overline{F}(x)} > \frac{1}{c}\right).$$

Let $x_{-1} = x_0 = 0$.

For fixed $t \geq k$, $J_F(x,t) \to \overline{F}(t) < c\overline{F}(t+1)$. So, for $r \in \mathbb{N}$, we can inductively define an unbounded, increasing sequence:

$$x_r = \inf(x : x - x_{r-1} \geq x_{r-1} - x_{r-2}, \sup_{y \geq x} J_F(y, k+r-1) \leq c\overline{F}(k+r), \sup_{y \geq x} K_F(y, k+r) \leq \overline{F}(k+r)).$$

Now define

$$h(x) = \begin{cases} kx/x_1 & x < x_1, \\ k + r - 1 + \frac{x - x_r}{x_{r+1} - x_r} & x \in [x_r, x_{r+1}), r \in \mathbb{N}. \end{cases}$$

For $x \in [x_r, x_{r+1}), r \in \mathbb{N}$, we have

$$k + r - 1 \leq h(x) < k + r.$$

So,

$$J_{F,h}(x) \leq J_F(x, k+r-1) \leq c\overline{F}(k+r) \leq c\overline{F}(h(x))$$

and

$$K_{F,h}(x) \leq K_F(x, k+r) \leq \overline{F}(k+r) \leq \overline{F}(h(x)).$$

By construction $h(x)$ is increasing, unbounded above and concave, which completes the proof. □

The next lemma is useful as a technical tool, and, among other uses, allows us to employ the dominated convergence theorem. It seems to be due to Kesten (see [11]). For a proof see, for instance, [32].

Lemma 2.2.2. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with common distribution F which is subexponential on the whole line. Then for any $\varepsilon > 0$ there exists a constant $K < \infty$ (dependent on ε) such that, for all $x \geq 0$ and $n \geq 1$,*

$$\mathbf{P}(X_1 + \cdots + X_n > x) \leq K(1+\varepsilon)^n \mathbf{P}(X_1 > x). \tag{2.17}$$

2.2.1 Conditions for Subexponentiality

Unlike long-tailedness, the condition for subexponentiality given in the Definition 2.2.1 is not usually easy to check. Several necessary or sufficient conditions have been

investigated, see, for example, [66, 78, 44, 69, 52, 63, 13]. Many of these depend on the properties of the hazard function and the hazard rate function, where, for a given distribution F, the hazard function is $Q(x) := -\log(\overline{F}(x))$, and, when the hazard function is differentiable, the hazard rate function is $q(x) := f(x)/\overline{F}(x)$. If F is subexponential then there is some function h such that F is h-insensitive. Thus a necessary condition is given by:

$$Q(x) - Q(x - h(x)) \to 0 \text{ as } x \to \infty.$$

However, as we have seen, it is not sufficient. A further necessary condition is given by the following proposition, also to be found in [63].:

Proposition 2.2.3. *Let F be a subexponential distribution with hazard function $Q(x)$ and h be a function such that F is h-insensitive. Then*

$$Q(x) - Q(x - h(x)) - Q(h(x)) \to -\infty,$$

as $x \to \infty$.

A sufficient condition is given by the following proposition, see, for example, [37] and references therein.

Proposition 2.2.4. *Let F be a long-tailed distribution on $\mathbb{R}$. Assume there exists $\gamma < 1$ and $A < \infty$ such that the hazard function $Q(x)$ satisfies*

$$Q(x) - Q(x - y) \leq \gamma Q(y) + A, \tag{2.18}$$

for all $x > 0$ and $y \in [0, x/2]$. If the function $e^{-(1-\gamma)Q(x)}$ is integrable over $\mathbb{R}^+$ then F is subexponential.
If $0 < \gamma < 1$ then the function $e^{-(1-\gamma)Q(x)}$ is integrable if F has a finite moment of order $\frac{1}{1+\gamma} + \varepsilon$ on $\mathbb{R}^+$ for some $\varepsilon > 0$.

If the hazard function $Q(x)$ is (eventually) concave then the following proposition is easier to check.

Proposition 2.2.5. *Let F be a long-tailed distribution on $\mathbb{R}$ and h be a positive non-decreasing continuous function such that F is h-insensitive. Let the hazard function $Q(x)$ be eventually concave, that is for $x \geq x_0$, for some $x_0 < \infty$. If*

$$x\overline{F}(h(x)) \to 0 \quad \text{as } x \to \infty, \tag{2.19}$$

then F is subexponential.

To prove this, we need the following technical lemma which was first used without proof in [28].

Lemma 2.2.3. *Let F be long-tailed. Then there exists a constant $C > 0$ such that for any $b > a > 0$,*

$$\int_a^b \overline{F}(x-y)F(dy) \leq C \int_a^b \overline{F}(x-y)\overline{F}(y)dy.$$

Proof. Let $y_0 = a$, $s = [b-a] + 1$ and $y_i = y_{i-1} + (b-a)/s, \quad i = 1, 2, \ldots, s$. Then $y_s = b$.

There exists a constant C such that for any $y > 0$, $\frac{\overline{F}(y)}{\overline{F}(y+1)} \leq \sqrt{C} < \infty$ since F is long-tailed.

Then

$$\begin{aligned}
\int_a^b \overline{F}(x-y)F(dy) &= \Sigma_{n=0}^{s-1} \int_{y_n}^{y_{n+1}} \overline{F}(x-y)F(dy) \\
&\leq \Sigma_{n=0}^{s-1} \int_{y_n}^{y_{n+1}} \overline{F}(x-y_{n+1})(\overline{F}(y_n) - \overline{F}(y_{n+1}))dy \\
&\leq \Sigma_{n=0}^{s-1} \int_{y_n}^{y_{n+1}} \sqrt{C}\overline{F}(x-y)\overline{F}(y_n)dy \\
&\leq \Sigma_{n=0}^{s-1} \int_{y_n}^{y_{n+1}} C\overline{F}(x-y)\overline{F}(y)dy \\
&= C \int_a^b \overline{F}(x-y)\overline{F}(y)dy.
\end{aligned}$$

□

Proof of Proposition 2.2.5. Without loss of generality we may assume that $x_0 = 0$. Since $Q(x)$ is concave, the minimum of the sum $Q(x-y) + Q(y)$ on the interval

$[h(x), x - h(x)]$ occurs at the endpoints of the interval. From the lemma, there exists a constant $C > 0$ such that

$$\begin{aligned}\int_{h(x)}^{x-h(x)} \overline{F}(x-y)F(dy) &\leq C \int_{h(x)}^{x-h(x)} \overline{F}(x-y)\overline{F}(y)dy \\ &= C \int_{h(x)}^{x-h(x)} \exp(-(Q(x-y)+Q(y)))dy \\ &\leq Cx \exp(-(Q(h(x))+Q(x-h(x)))) \\ &= Cx\overline{F}(h(x))\overline{F}(x-h(x)),\end{aligned}$$

and so

$$\begin{aligned}\int_{h(x)}^{x-h(x)} \frac{\overline{F}(x-y)}{\overline{F}(x)}F(dy) &\leq Cx\overline{F}(h(x))\frac{\overline{F}(x-h(x))}{\overline{F}(x)} \\ &= o(1).\end{aligned}$$

Therefore Theorem 2.2.1 shows that F is subexponential. □

As an example, consider the Weibull distribution with shape parameter $0 < \beta < 1$ and tail distribution function given by $\overline{F}(x) = e^{-x^\beta}$. Then $Q(x) = x^\beta$ is concave for $\beta \in (0, 1)$. and $xe^{-x^\beta} \to 0$ as $x \to \infty$. Hence, as previously stated, F is subexponential.

2.2.2 Closure properties for subexponential distributions

Closure properties of regularly varying distributions are well-known. In some sense the subexponential distributions are the natural extension of the regularly varying distributions. The class of regularly varying distributions is closed under convolutions (see [16]). Also in [31] it is shown that if F is a regularly varying distribution and G is a distribution such that $\overline{G} = o(\overline{F})$ then the product convolution (that is, if X is a non-negative random variable with distribution F and Y is an independent non-negative random variable with distribution G, then the product convolution is the distribution of XY) is regularly varying. The situation is not so straightforward when considering subexponential distributions. In the case of the ordinary convolution for a

sum, it is known that in general the class of subexponential distributions is not closed under convolution (see, for example [57]).
First we consider closure under tail equivalence relations. Proofs of these results may be found in, for example, [37] and references therein

Proposition 2.2.6. *Let F be subexponential on the whole line, and G be long-tailed. Let F and G be weakly tail equivalent, $\overline{F} \asymp \overline{G}$. Then G is subexponential on the whole line.*

Proposition 2.2.7. *Let F be subexponential on the whole line. Let F and G be proportionally tail equivalent, that is there exists a constant $c > 0$ such that $\overline{F}(x) \sim c\overline{G}(x)$. Then G is subexponential on the whole line.*

Now we consider some results related to convolutions. The next two propositions are taken from [31] and give some sufficient conditions for a convolution to be subexponential.

Proposition 2.2.8. *Let F and G be subexponential on the whole line. If F and G are weakly tail equivalent then $F * G$ is subexponential on the whole line.*

Proposition 2.2.9. *Let F be subexponential on the whole line. If $\overline{G}(x) = o(\overline{F}(x))$ then $\overline{F * G}(x) \sim \overline{F}(x)$. Hence $F * G$ is also subexponential on the whole line.*

Proposition 2.2.10. *Let F be a (reference) subexponential distribution on the whole line. Let $G_1, \ldots, G_n$ be distributions such that, for each $i = 1, \ldots, n$, $\overline{G}_i(x) \sim c_i\overline{F}(x)$ for some constants $c_i \geq 0$. Then $\overline{G_1 * \ldots * G_n}(x) \sim (c_1 + \cdots + c_n)\overline{F}(x)$.*
If $c_1 + \cdots + c_n > 0$ then $G_1 * \ldots * G_n$ is subexponential.

The study of subexponential products begins with [21] and the following two propositions, which we have translated into the language of h-insensitivity.
In the next two propositions let X and Y be independent positive random variables with distributions F and G and let H be the distribution of the product XY.

Proposition 2.2.11. *Let F be subexponential. If there exists a function positive, continuous, non-decreasing functions h tending to ∞ for which F is h-insensitive, such that i) $h(x) = o(x)$ and ii) $\overline{G}(h(x)) = o(\overline{H}(x))$ then H is subexponential. A sufficient condition that we may require instead of ii) is that $\overline{G}(h(bx)) = o(\overline{F}(x))$ for some $b > 0$.*

Proposition 2.2.12. *Let F be subexponential and Y be bounded (above). Then H is subexponential.*

2.3 The Boundary Class

In certain applications, given a long-tailed or subexponential distribution F, it is necessary to know that there exists a positive non-decreasing function h tending to ∞ for which F is h-insensitive, and for instance the quantity defined in (2.14) or the quantity $\overline{F}(h(x))$ (or both) tends to zero quickly enough. It may be difficult to find a suitable function h, and the fact that trial functions have not worked does not, in general, mean that no such h-function exists.

These quantities have the property that, given one such h-function h_1 for which the quantity of interest converges to zero quickly enough, then for any other (positive non-decreasing tending to ∞) function h_2 such that eventually $h_2(x) > h_1(x)$ then the quantity of interest with h_1 replaced by h_2 will also decay quickly enough. We shall refer to this as *increasing function* behaviour.

However, if we arbitrarily increase the h-function there will come a point when F is no longer h-insensitive. This suggests that there should be some boundary to the class of h-functions for which F is h-insensitive. We refer to this as the boundary class for F. For any function H in the boundary class the distribution is not H-insensitive, but F is h-insensitive for any function h that satisfies $h(x) = o(H(x))$.

It is not the case that the boundary class, as we define it, always exists (for instance, long-tailed distribution functions that are slowly varying (see the appendix for the definition of a slowly varying function, and other related concepts)). However, all

commonly encountered long-tailed distributions that are not slowly-varying do possess a boundary class. We shall show that all functions in the boundary class are weakly equivalent, and that this means that the boundary class can be generated by a single function $H(x)$, and all multiples of $H(x)$.

We shall show that if the quantity of interest decays quickly enough for all multiples of a generator $H(x)$ then there exists some h-function $h(x)$ for which the quantity decays quickly enough and for which F is h-insensitive. A generator for the boundary class is usually easy to find, and almost trivial for absolutely continuous distributions, whereas we may not easily be able to find the suitable h-function.

2.3.1 Definition and Properties

Given a long-tailed function F we want to consider the class of positive, continuous, non-decreasing functions h tending to ∞ and with $h(x) < x/2$ for which F is h-insensitive. In order to avoid repetition we refer to this class of functions as $\hbar_F$. We wish to describe the upper boundary of this class, when it exists.

Definition 2.3.1. *Let F be a long-tailed distribution. The* boundary class *(for F), $\mathcal{H}_F$, consists of all continuous, non-decreasing functions $H(x)$ such that $h(x) \in \hbar_F$ if and only if $h(x) = o(H(x))$.*

Remark 2.3.1. In most cases the boundary class for a long-tailed distribution does exist, however we note that any slowly varying function $\overline{F}$ does not possess a boundary class as all non-decreasing functions $h(x)$ defined on the positive reals such that $0 < h(x) < x/2$ and $h(x) \to \infty$ as $x \to \infty$ is such that F is h-insensitive.

We examine the structure of the boundary class $\mathcal{H}_F$, and show that all functions in $\mathcal{H}_F$ are weakly tail equivalent.

Proposition 2.3.1. *Let F be a long-tailed distribution. Let $H_1(x)$ belong to the boundary class $\mathcal{H}_F$. Then $H_2(x) \in \mathcal{H}_F$ if and only if $H_2(x) \asymp H_1(x)$.*

Proof. Clearly, if $H_2(x) \asymp H_1(x)$ then $H_2(x) \in \mathcal{H}_F$. So, consider a function $H_2(x)$ for which $\liminf \frac{H_2(x)}{H_1(x)} = 0$. We shall construct a function $h_1(x) \in \hbar_F$ which is not

$o(H_2(x))$. There exists a sequence, tending to infinity, $0 = x_0 < x_1 < \ldots$ with $\varepsilon_n := \frac{H_2(x_n)}{H_1(x_n)}$ such that $\lim_{n\to\infty} \varepsilon_n = 0$.

By the continuity of $H_1(x)$ and $H_2(x)$ we can find a sequence of points $0 < y_1 < x_1 < y_2 \ldots$ such that $0 < \frac{H_2(x)}{H_1(x)} < 2\varepsilon_n$ for all $y_n \leq x < x_n$.

For $n \geq 1$, define

$$h_1(x) = \begin{cases} H_2(x_{n-1}) & \text{for } x \in [x_{n-1}, y_n), \\ \frac{x_n - x}{x_n - y_n} H_2(x_{n-1}) + \frac{x - y_n}{x_n - y_n} H_2(x_n) & \text{for } x \in [y_n, x_n). \end{cases}$$

Now, $h_1(x) = o(H_1(x)$ by construction, so $h_1(x)$ satisfies the long-tail property, but $\liminf \frac{h_1(x)}{H_2(x)} = 1$, so $H_2(x) \notin \mathcal{H}_F$. We can clearly repeat this argument if $\liminf \frac{H_1(x)}{H_2(x)} = 0$. Hence, if $H_2(x) \in \mathcal{H}_F$ then $H_2(x) \asymp H_1(x)$. □

As we remarked in the introduction to this chapter, we are interested in using the boundary class to demonstrate the existence of a function $h \in \hbar_F$ satisfying certain properties. We refer to these properties as *increasing function properties*. Let $\Pi_h(x)$ be a quantity, dependent on choice of a function h. Then an increasing function property is a property of the form

$$\lim_{x\to\infty} \Pi_h(x) = 0, \tag{2.20}$$

such that if $h_2(x) \geq h_1(x)$ for all $x \geq 0$ and $\lim_{x\to\infty} \Pi_{h_1}(x) = 0$ then $\lim_{x\to\infty} \Pi_{h_2}(x) = 0$.

The next proposition will show that if an increasing function property is satisfied by *all* functions in the boundary class $\mathcal{H}_F$, then it is satisfied by at least one $h \in \hbar_F$. Since all functions in $\mathcal{H}_F$ are weakly-tail equivalent, it will then be sufficient to show that the condition holds for all multiples $\{cH(x); c \in \mathbb{R}^+, cH(x) < x/2\}$ of any particular function $H \in \mathcal{H}_F$. We shall then say that $H(x)$ generates the boundary class $\mathcal{H}_F$.

Proposition 2.3.2. *Let F be a distribution function possessing a boundary class $\mathcal{H}_F$. Then there exists some function $h(x) \in \hbar_F$ satisfying $\lim_{x\to\infty} \Pi_h(x) = 0$ if and only if $\lim_{x\to\infty} \Pi_{cH}(x) = 0$ for every $cH(x)$, where $c > 0$ and $cH(x) < x/2$, where $H(x)$ is any generator of $\mathcal{H}_F$.*

Proof. Choose any $H(x) \in \mathcal{H}_F$, and let $c_n = 2^{-n}, \quad n \in \mathbb{N}$.

Define an infinite sequence $0 = y_1 < x_1 < y_2 < \ldots$ recursively, for $r \in \mathbb{N}$, by

$$y_1 = 0;$$

$$x_r = \max(y_r + 1, \sup_{x>0}\{x : \Pi_{c_r H}(x) > c_r\});$$

$$y_{r+1} = \inf_{x>x_r+1}\{x : H(x) = 2H(x_r)\}.$$

By construction this sequence tends to infinity.

For $x \geq 0$, define

$$h(x) = \begin{cases} c_r H(x) & \text{for } x \in [y_r, x_r), \\ c_r H(x_r) & \text{for } x \in [x_r, y_{r+1}). \end{cases}$$

Hence, if the increasing function property holds for all (sufficiently small) multiples of $H(x)$, then it holds for $h(x)$, which, by construction, is $o(H(x))$. Conversely, if the increasing function property holds for some $h(x)$, then it holds for any function $g(x)$ such that $h(x) = o(g(x))$, and hence for all functions in class $\mathcal{H}_F$. □

Remark 2.3.2. Let F be a distribution function such that $\overline{F}(x) = f_1(x)f_2(x)$, where each of f_1 and f_2 are long-tailed. Let the boundary class for f_i be $\mathcal{H}_i$, $i = 1, 2$ and generated by $H_i(x)$ respectively. Assume that $H_2(x) = o(H_1(x))$. Then the boundary class for F is $\mathcal{H}_2$.

Remark 2.3.3. Very heavy distributions that are slowly varying do not always possess a boundary class. But slowly varying functions are always x-insensitive. We are interested in applications in finding functions $h \in \hbar_F$ such that are $h(x) \leq x/2$, and so the boundary of the slowly varying functions is not a significant factor. In particular, if F is a distribution function such that $\overline{F}(x) = f_1(x)f_2(x)$, where f_1 is slowly varying, f_2 is not slowly varying and f_2 has boundary class $\mathcal{H}$ then F has boundary class $\mathcal{H}$.

Proposition 2.3.3. *Let F be an absolutely continuous long-tailed distribution function with continuous strictly positive density $f(x)$ and hazard rate $q(x) = \frac{f(x)}{\overline{F}(x)}$. Let $H(x) = 1/q(x)$. Then the boundary class of F is generated by $\{cH(x); c \in \mathbb{R}^+, cH(x) < x/2\}$.*

Proof. Since F is long-tailed then for any $h \in \hbar_F$ we know that $\overline{F}(x - h(x)) = \overline{F}(x) + o(\overline{F}(x))$. Hence, $x - h(x) = \overline{F}^{-1}(\overline{F}(x)(1 + o(1)))$, and since $\overline{F}^{-1}$ has a derivative at all points in its domain, $h(x) = o(\overline{F}(x)(-\overline{F}^{-1})'(\overline{F}(x)))$, where the negative sign has been introduced to make the function inside the little-o positive. However, $\overline{F}(x)(-\overline{F}^{-1})'(\overline{F}(x)) = 1/q(x)$.
Conversely, if $h(x) = o(1/q(x))$ then it is easy to show that F is h-insensitive. □

We give some examples of calculating the Boundary class.

i) Let $F \in \mathcal{R}_{-\alpha}$, that is F is regularly varying (at infinity) with index $-\alpha < 0$ and $\overline{F}(x) = l(x)x^{-\alpha}$, $x > 1$, where $l(x)$ is slowly varying (but see Appendix for definitions). The slowly varying function $l(x)$ is x-insensitive, so by the remark above we need only consider the boundary class for $f_2(x) = x^{-\alpha}$. For f_2 we have $q_2(x) = \alpha/x$. Hence the boundary class is generated by $H(x) = x$.

ii) Let $\overline{F}(x) = \exp(-\gamma x^{\beta})$, $x > 0$, where $0 < \beta < 1$. Then $q(x) = \gamma\beta x^{-1+\beta}$, and the boundary class is generated by $H(x) = x^{1-\beta}$.

iii) Let $\overline{F}(x) = f_1(x)\exp(-\gamma(\log(x))^{\alpha}) := f_1(x)f_2(x)$, $x > 1$, where $\alpha \geq 1$ and $f_2(x) = o(f_1(x))$. We note that this class of functions includes regular variation as defined above, and also is tail equivalent to log-normal for $\alpha = 2$. We need only consider $f_2(x)$. We then find that the boundary class is generated by $H(x) = x(\log(x))^{1-\alpha}$.

As an example of the application of these ideas, consider the problem of demonstrating that the Weibull distribution with shape parameter $0 < \beta < 1$ and tail distribution function given by $\overline{F}(x) = e^{-x^{\beta}}$ is subexponential. We shall use Proposition 2.2.5. The hazard function $Q(x) = x^{\beta}$ is concave for $\beta \in (0, 1)$. Hence we need to demonstrate that, for some function $h \in \hbar_F$, $x\overline{F}(h(x)) = o(1)$. This is an increasing function

property, so instead of having to find a specific h, we may choose a function in the boundary class $\mathcal{H}_F$, say $H(x) = x^{1-\beta}$ and check that for all $c > 0$,

$$xe^{-\beta(cx^{1-\beta})} = o(1),$$

which is clearly true. Hence, as previously stated, F is subexponential.

2.3.2 The Boundary Class and Auxiliary Functions

Our concept of the boundary class is also closely related to the concept of an auxiliary function, introduced by de Haan [26], see also, for example, [9, 32, 71]. However, the boundary class can exist when there is no auxiliary function, or when the conditions of the previous proposition are not met, and hence it is a more general concept.

The concept of an auxiliary function was introduced to characterize distributions that lie in the maximum domain of attraction of the Gumbel extreme value distribution, $MDA(\Lambda)$. A positive function $a(x)$ is an auxiliary function for the distribution function $F(x) \in MDA(\Lambda)$ (with $F(x) < 1$ for all $x \in \mathscr{R}$) if and only if

$$\lim_{x\to\infty} \frac{\overline{F}(x + ta(x))}{\overline{F}(x)} = e^{-t}, \text{ for all } t \in \mathbb{R}. \tag{2.21}$$

If an auxiliary function exists, all such functions are asymptotically equivalent, possible choices are the reciprocal of the hazard rate and the mean excess function, see, for example, [32], and any auxiliary function is in the boundary class.

The concept of the auxiliary function may be extended. For instance, we consider regularly varying $\overline{F}(x) \in \mathcal{R}_{-\alpha}$ (see Section 6.2 for definition) we have an auxiliary function $a(x)$, which may be taken to be $a(x) = x$, satisfying

$$\lim_{x\to\infty} \frac{\overline{F}(x + ta(x))}{\overline{F}(x)} = (1 + t)^{-\alpha}, \text{ for all } t \in \mathbb{R}, \tag{2.22}$$

and again any auxiliary function is in the boundary class.

The concept of the boundary class is more general than that of the auxiliary function. We can construct a (subexponential) distribution function belonging to the class of

intermediately regularly varying distributions. Let $c_1(x) = 2 + \sin(\log x)$ and $\overline{F}(x) = \max(1, c_1(x)x^{-\alpha}),\ \alpha > 0$. Then F is in the intermediately varying class. Indeed,

$$\frac{\overline{F}(\lambda x)}{\overline{F}(x)} = \frac{2 + \sin(\log x + \log \lambda)}{2 + \sin(\log x)} \lambda^{-\alpha},$$

so

$$\lim_{\lambda \downarrow 1} \liminf_{x \to \infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = 1.$$

It is straightforward to check that this has boundary class generated by $H(x) = x$ but the limit in (2.22) does not exist. Indeed, if $h(x) = o(x)$ then $c_1(x + h(x)) \sim c_1(x)$, but

$$\frac{c_1(x + cH(x))}{c_1(x)} = \frac{2 + \sin(\log((1 + c)x))}{2 + \sin(\log(x))},$$

which does not possess a limit as $x \to \infty$.

We can also construct a similar example for what we might call an intermediate Weibull distribution. Let $c_2(x) = 2 + \sin(x^\beta)$ and define

$$\overline{F}(x) = \max(1, c_2(x)e^{-x^\beta}),$$

where $0 < \beta < 1$.

It is straightforward to show that F is long-tailed with boundary class generated by $H(x) = x^{1-\beta}$. To show that F is subexponential on $\mathbb{R}^+$, we consider the hazard function of F,

$$Q(x) = x^\beta - \log(2 + \sin(x^\beta)).$$

The hazard function is not eventually concave, so we turn to Proposition 2.2.4. We note that

$$x^\beta - \log 3 \leq Q(x) \leq x^\beta$$

for all $x > 0$. Hence, for $y \in [0, x/2]$,

$$\begin{aligned} Q(x) - Q(x - y) \leq & x^\beta - (x - y)^\beta + \log 3 \\ \leq & \beta y^\beta + \log 3 \\ \leq & \beta Q(y) + 2 \log 3, \end{aligned}$$

which satisfies condition (2.18). Also F possesses all finite moments, and so the conditions of Proposition 2.2.4 are met and F is subexponential.

However, if the auxiliary function exists it can be taken to be $H(x)$, but

$$\frac{c_1(x+tH(x))}{c_1(x)} = \frac{2+\sin(x^\beta+t+o(1))}{2+\sin(x^\beta)}$$

does not possess a limit as $x \to \infty$, so there is no auxiliary function for F.

In both the previous examples Proposition 2.3.3 can be used to find the boundary class because of the smoothness of the functions $c_1(x)$ and $c_2(x)$. However, it would be easy to replace these functions with functions having the same oscillatory behaviour but without the smoothness, This would not affect the boundary class, but would no longer allow the boundary class to be generated by the reciprocal of the hazard rate.

2.3.3 Classifying Distributions

The boundary class gives us a natural way of classifying some long-tailed distributions by considering which distributions possess a particular boundary class. In [37] it is shown (Theorem 2.46) that a distribution F is intermediate regularly varying if and only if F is $o(x)$-insensitive. This class includes not only the distribution functions that have behaviour that is similar to regular variation but also the slowly varying distribution functions. A trivial alteration to the proof in [37] gives us

Proposition 2.3.4. *A distribution on $\mathbb{R}$ is intermediate regularly varying but not slowly varying if and only if F possesses a boundary class which is generated by $H(x) = x$.*

The class of intermediate regularly varying distributions that are not slowly varying are the generalisations of regularly varying distributions (and hence the general Pareto distributions) that are of most use in applications.

Other classes remain to be investigated. It is to be expected, for instance, that the class of distributions that have boundary class generated by $H(x) = e^{-x^\beta}$ for $0 < \beta < 1$ should form the natural generalisations of the long-tailed Weibull distributions. The exact structure of this and other related classes remains an open question.

Chapter 3

Bounds for Geometric Sums of Heavy-tailed Random Variables

This chapter is concerned with estimating the tail distribution of a geometric sum of independent non-negative random variables which have common subexponential distribution function F. By a *geometric sum* we mean the following. Let $X_1, X_2, \ldots$ be a sequence of non-negative independent and identically distributed random variables with unbounded support on the positive half-line. Let ν be an independent counting random variable with geometric distribution,

$$\mathbf{P}(\nu = k) = p(1-p)^{k-1},\ k \geq 1,\ 0 < p < 1.$$

Then

$$S_\nu = \sum_{i=1}^{\nu} X_i \tag{3.1}$$

is a geometric sum. We define $S_0 = 0$.

First we shall describe why the geometric sum is an object of interest and give some examples of areas of application. We shall then discuss the asymptotics of the tail distribution and show that if the asymptotics are used as an approximation to the tail distribution that this may result in huge errors. We shall review some methods that have been used to obtain bounds for the accuracy of the asymptotic approximation. We shall then derive new bounds for the error. These new bounds are based on a

method developed by Kalashnikov and Tsitsiashvili in [50, 51]. We shall then show how to use our results to obtain bounds for some specific examples, and compare the accuracy of the bounds with the actual tail derived by numerical computation.

3.1 Geometric Sums

Typically geometric sums arise in situations when the phenomenon of interest has a cyclic nature. As an example consider this simple problem from reliability theory. A system has a vital component with built in redundancy, that is there are N copies of the component, and only failure of all N components causes the system to fail. When one of the components fails it is immediately removed to be repaired and one of the other (working) components replaces it (if one is still available). There is one repairing unit where repair of the failed component begins immediately the repair unit becomes idle. If the repair unit is busy repairing a previously failed component then failed components form a first-come-first-served queue. The lifetimes of the components from repair to failure are i.i.d. non-negative random variables with common distribution F. The repair time is a non-negative random variable, independent of everything else with distribution G. Let $N(t)$ be the number of components out of service at time t. The cyclic nature comes from considering either the moments at which $N(t)$ drops from 1 to 0, or alternatively climbs from 0 to 1. This regenerative structure repeats itself unless $N(t)$ reaches N, when the whole system fails. In any cycle the probability of this event occurring is the same, say p. In a cycle in which $N(t) = n$ (a 'bad' cycle), let τ be the random variable measuring how long after the beginning of that cycle $N(t)$ reaches level N. In a 'good' cycle, that is conditional on $N(t)$ not reaching N (which happens with probability $1 - p$), the times between the start of that cycle and the next are i.i.d. non-negative random variables $X_1, X_2, \ldots$. Then the time to failure starting from the beginning of a particular cycle is given by

$$T = \sum_{i=1}^{\nu-1} X_i + \tau,$$

where ν is a geometric random variable, independent of the X_i's, with parameter p and τ is independent of everything else (and we have used the convention that the sum is 0 if $\nu = 1$). Clearly many situations in reliability theory fall into this framework, see, for instance, [23, 43].

Another major class of applications comes from situations which can be modelled as the maximum of a random walk with negative drift. We have already mentioned in Chapter 1, see equation (1.7), the ruin problem of the Sparre-Andersen model, and also the calculation of the tail distribution of the stationary waiting time of a first-come-first-served single server queue with general independent interarrival times and general independent service times (GI/GI/1), see, for instance [5]. An excellent overview that provides additional examples is [49]. The connection with geometric sums is due to Feller [34].

We consider a random walk with negative drift $\sigma_0 = 0$, $\sigma_n = \xi_1 + \cdots + \xi_n$, for $n \geq 1$, where the sequence of random variables $\xi_1, \xi_2, \ldots$ are independent and identically distributed and $\mathbf{E}(\xi_1) < 0$. We want to find the distribution of the maximum of the random walk

$$M = \sup_{0 \leq k < \infty} \sigma_k. \tag{3.2}$$

It is straightforward from the strong law of large numbers that M is almost surely finite.

Feller introduced the random variable representing the *first strict ascending ladder epoch* (subsequently we shall abbreviate this to the ladder epoch when there is no chance of confusion),

$$\tau_1 = \inf\{k : \sigma_k > 0, k \geq 1\}, \tag{3.3}$$

(with the convention that $\inf \emptyset = \infty$) and the corresponding (first strict ascending) ladder height

$$L_1 = \sigma_{\tau_1}. \tag{3.4}$$

with the convention that $\sigma_\infty = -\infty$. Because M is almost surely finite τ_1 and L_1 are

defective random variables and

$$p = \mathbf{P}(\tau_1 = \infty) > 0. \tag{3.5}$$

We can then define recursively

$$\tau_n = \begin{cases} \inf\{k : \sigma_k > L_{n-1}\} & \tau_{n-1} < \infty, \\ \infty & \tau_{n-1} = \infty, \end{cases} \tag{3.6}$$

and

$$L_n = \sigma_{\tau_n}, \tag{3.7}$$

for $n \geq 2$, and let $L_0 = 0$. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common distribution F such that

$$\begin{aligned} \mathbf{P}(X_n \leq x) &\stackrel{d}{=} \mathbf{P}(L_n - L_{n-1} \leq x | L_n > -\infty) \\ &\stackrel{d}{=} \mathbf{P}(L_1 \leq x | \tau_1 < \infty), \end{aligned}$$

for $n \geq 1$ (and where $\stackrel{d}{=}$ stands for equality in distribution). Also, the conditional probability that $\tau_n = \infty$ given that $\tau_{n-1} < \infty$

$$\mathbf{P}(\tau_n = \infty | \tau_{n-1} < \infty) = p.$$

Let $\nu = \min\{n : \tau_n = \infty, ; n \geq 1\}$. Then ν is a geometric random variable with parameter p. Now the maximum of the random walk can be written as

$$\begin{aligned} M &= L_1 + \cdots + L_{\nu-1} \\ &\stackrel{d}{=} X_1 + \cdots + X_{\nu-1}. \end{aligned}$$

We have

$$\mathbf{P}(M = 0) = p$$

Also, conditional on $\nu > 1$, if $\nu*$ is an independent geometric r.v. with parameter p then

$$X_1 + \cdots X_{\nu-1} \stackrel{d}{=} X_1 + \cdots + X_{\nu*}, \tag{3.8}$$

so

$$\mathbf{P}(M \leq x) = p + (1-p)\mathbf{P}(X_1 + \cdots + X_\nu \leq x), \tag{3.9}$$

and we have written the distribution of a random walk with negative drift in terms of a geometric sum. It is important to realise that the geometric sum is not written in terms of random variables with the same distribution as the original steps in the random walk, ξ_i. Rather the random variables X_i have the conditional ascending ladder height distribution F. In general explicit formulas for p, the parameter of the geometric distribution of ν, and F do not exist. In some situations, for instance the classical risk model (1.4) where the renewal process is a Poisson process, there are explicit representations for p and F (see, for example, [47]). However there are many works which give approximations and bounds for p and F (for example [20, 62, 15]). From now on we concentrate on the geometric sum S_ν in the case when F is subexponential, but for alternative approaches to studying the maximum of a random walk in the subexponential case we refer the reader to [66, 80] and the pleasingly probabilistic approach in [82].

3.2 Asymptotic tail of the Geometric Sum with Subexponential Summands

The asymptotic tail of the distribution is well-known, and can be traced back to [79] and was first fully developed in [33]. We have

$$\begin{aligned}
\mathbf{P}(S_\nu > x) =& \mathbf{E}(\mathbf{P}(X_1 + \cdots + X_\nu | \nu)) \\
\sim& \mathbf{E}(\nu \mathbf{P}(X_1 > x)) \qquad (3.10) \\
=& \mathbf{E}(\nu)\mathbf{P}(X_1 > x)) \\
=& \frac{1}{p}\overline{F}(x), \qquad (3.11)
\end{aligned}$$

where the interchange of expectation and limit in (3.10) is justified by the use of Kesten's lemma 2.2.2 and the dominated convergence theorem. We note that the

assumption of independence between ν and the random variables $X_1, X_2, \ldots$ can be weakened by strengthening the conditions on the distribution F, see [25].
In general the accuracy of this asymptotic relation (3.11) is extremely bad even for reasonably large values of x.

Example 3.2.1. For instance, if we take F to be a shifted Pareto distribution with shape parameter $\alpha = 5$ and take $p = 0.2$, so that

$$\overline{F}(x) = \begin{cases} 1 & x \leq 1, \\ x^{-5} & x > 1 \end{cases}$$

then we have $\mathbf{E}(X_1) = 1.25$ and $\mathbf{V}(X_1) = 5/48$.
Numerically evaluating the tail of the geometric sum using a discretized Panjer recursion (see [68, 73]) and comparing with (3.11),

$$\mathbf{P}(S_\nu > 30) = 0.00547,$$
$$\mathbf{E}(\nu)\mathbf{P}(X > 30) = 2.06 \times 10^{-7}.$$

This gives us a relative error of more than 26000 in using the asymptotic expression at $x = 30$.
A better overall picture is obtained by plotting the logarithm (to base 10) of the relative error defined by

$$\Delta_F(x) := \frac{\mathbf{P}(S_\nu > x) - \mathbf{E}\nu\overline{F}(x)}{\mathbf{E}\nu\overline{F}(x)}, \tag{3.12}$$

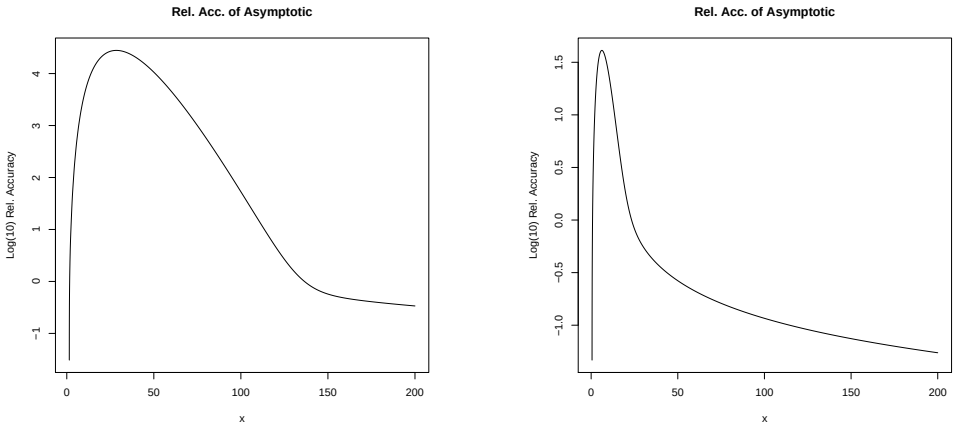
for different values of the argument. The logarithm of the relative error for the example above is in Figure 3.2.1.

Geometric sums with the same geometric parameter p and precisely the same tail behaviour for F can have wildly different relative error.

Example 3.2.2. Consider instead the geometric sum with F a Pareto distribution with shape parameter $\alpha = 5$ and take $p = 0.2$, and

$$\overline{F}(x) = \begin{cases} 1 & x \leq 0, \\ (1+x)^{-5} & x > 0. \end{cases}$$

The plot of the logarithm of the relative error is in Figure 3.2.2. Note in particular the difference in the scale on the y-axis.



(a) Example 3.2.1 (b) Example 3.2.2

Figure 3.1: Logarithm of Relative Error for two Geometric sums with Tail-equivalent Increments

Hence the need to establish bounds for the relative accuracy of the asymptotic relation. Of particular importance is the upper bound for the tail of the distribution, which gives us the worst case scenario. The need for such bounds was understood very early on in the classical setting, as the famous Lundberg bound [58]. In the heavy-tailed setting the accuracy of the asymptotic expression is much less well understood.

3.3 Previous Approaches

There have been many studies devoted to bounds for ruin probabilities and geometric sums in the presence of heavy tails. We mention some of them, mostly in the context of calculation of ruin probabilities for the Sparre-Andersen model. Willekens and Teugels [81] consider higher order expansions of (3.11). In [7] methods for simulation of ruin probabilities are considered. In [45, 42, 30] algorithms are developed to obtain both upper and lower bounds for the ruin probability. In the context of risk models

again, some methods replace the actual claim distribution with another distribution in such a way that the first few moments coincide, and so that the ruin probability for the new process is easier to determine. Among such approaches are [27, 68]. Methods for numerical inversion of the Laplace transform of the ruin function are considered in [2]. An analytical approach to the general problem of geometric sums is to be found in [49].

3.4 Upper Bound for the Geometric Sum

Our approach follows that of Kalashnikov and Tsitsiashvili in [50, 51]. It differs in that our approach is, as far as possible, probabilistic in nature and makes use of the h-insensitivity of a subexponential distribution. Also, Kalashnikov and Tsitsiashvili ignored the fact that the distribution of S_ν has positive mass p at $x = 0$ which means that their formulas are incorrect. In addition, once we have derived an expression for the upper bound the way we implement it in practical situations is different.

As a reminder, a geometric sum is given by $S_\nu = \sum_{i=1}^{\nu} X_i$ where $X_1, X_2, \ldots$ are i.i.d. non-negative random variables with subexponential distribution and ν is an independent counting random variable with geometric distribution with parameter p. We shall consider the relative error in using the asymptotic expression (3.11) as an approximation for S_ν. We recall the relative error in equation (3.12), and will also find it convenient to define

$$\Delta[a, b] = \sup_{a \le x \le b} \Delta(x).$$

We note that, since (3.11) gives the correct asymptotics then

$$\lim_{x \to \infty} \Delta_F(x) = 0.$$

We choose a positive non-decreasing concave function $h(x) < x/2$ which tends to ∞, $h \in \hbar_F$, so that F is h-insensitive, which means that

$$\overline{F}(x - h(x)) \sim \overline{F}(x).$$

Consider the total probability formula

$$\mathbf{P}(S_\nu > x) = \mathbf{P}(X_1 > x)\mathbf{P}(\nu = 1) + \mathbf{P}(S_\nu > x|\nu > 1)\mathbf{P}(\nu > 1). \tag{3.13}$$

Conditional on $\nu = 1$,

$$S_\nu \stackrel{d}{=} X_1.$$

Conditional on $\nu > 1$, we argue similarly to the way we did in (3.8) by introducing independent copies of X_1 and ν which we call X and $\nu*$; we then obtain

$$S_\nu \stackrel{d}{=} X + S_{\nu*}.$$

Concentrating first on $\mathbf{P}(S_\nu > x|\nu > 1) = \mathbf{P}(X + S_\nu > x)$, we partition using the size of X relative to $h(x)$ and obtain the bound

$$\begin{aligned}\mathbf{P}(S_\nu > x|\nu > 1) \leq &\mathbf{P}(X \leq h(x), S_\nu > x - h(x)) \\ &+ \mathbf{P}(h(x) < X_1 \leq x - h(x), S_\nu > x - X) \\ &+ \mathbf{P}(X > x - h(x)).\end{aligned} \tag{3.14}$$

From the definition of $\Delta_F(x)$ (3.12) we have

$$\mathbf{P}(S_\nu > x) = \frac{1}{p}(1 + \Delta_F(x))\overline{F}(x). \tag{3.15}$$

Substituting (3.15) into (3.14) we can then bound the conditional probability

$$\begin{aligned}&\mathbf{P}(S_\nu > x|\nu > 1) \\ \leq &\mathbf{P}(X \leq h(x))\mathbf{P}(S_\nu > x - h(x)) \\ &+ \int_{h(x)}^{x-h(x)} \mathbf{P}(S_\nu > x - y)\mathbf{P}(X \in dy) + \mathbf{P}(X > x - h(x)) \qquad (3.16) \\ \leq &\frac{1}{p}(1 + \Delta_F(x - h(x)))\overline{F}(x - h(x))F(h(x)) \\ &+ \frac{1}{p}(1 + \Delta_F[h(x), x - h(x)]) \int_{h(x)}^{x-h(x)} \overline{F}(x - y)F(dy) + \overline{F}(x - h(x)). \qquad (3.17)\end{aligned}$$

On examining (3.17) we see that we have quantities related to (2.13),(2.14).
We now turn to the construction of the upper bound for $\Delta_F(x)$.

Substituting (2.13) and (2.14) into (3.17) and then into (3.13) we obtain

$$\begin{aligned}\frac{\mathbf{P}(S_\nu > x)}{\overline{F}(x)} \leq & \frac{1-p}{p}(1+\Delta_F(x-h(x)))(K_{F,h}(x)+1)F(h(x)) \\ & + \frac{1-p}{p}(1+\Delta_F[h(x),x-h(x)])J_{F,h}(x) \\ & + (1-p)(K_{F,h}(x)+1)+p, \end{aligned} \tag{3.18}$$

We expect the upper bound for $\Delta(x)$ to be a monotonic function $\Delta_F^+(x)$, decreasing to 0, such that

$$\Delta_F(x) \leq \Delta_F^+(x) \quad \text{for all } x > 0. \tag{3.19}$$

The upper bound will be of the form

$$\Delta_F^+(x) = Cg(x) \tag{3.20}$$

for some suitably chosen test function g and constant $0 < C < \infty$. It is clear that, in order for Δ_F^+ to have the desired properties then g must be a positive monotonic function decreasing to 0.

With this in mind we choose g (the choice of g depending upon our prior choice of h) so that

$$g(x) \downarrow 0; \tag{3.21}$$

$$\max(J_{F,h}(x), K_{F,h}(x), \overline{F}(h(x))) = O(g(x)); \tag{3.22}$$

$$g(x-h(x)) \sim g(x). \tag{3.23}$$

The existence of such a function g satisfying (3.22) is guaranteed by Lemma 2.2.1, since we may choose $g(x) = \overline{F}(h(x))$, with $h(x)$ concave, in Lemma 2.2.1. The concavity of h then implies $h(x-h(x)) \geq h(x) - h(h(x))$ which in turn gives us (3.23).

The idea of the test function g is that it should in some sense be close to Δ_F. Define

$$C(x) = \max\left(0, \frac{\Delta(x)}{g(x)}\right), \tag{3.24}$$

$$C[a,b] = \max_{a \leq x \leq b} C(x) \tag{3.25}$$

as measures of the closeness of g to its target Δ_F. If we can bound $C(x)$ above for all $x > 0$ by some constant C then we will have (3.20) as desired.

Although $\Delta_F(x)$ may be positive or negative (or zero), we note that all the terms in the expression (3.17) are positive. Substituting this into (3.24) we obtain

$$C(x) \leq (f_1(x) + f_2(x))C[h(x), x] + f_3(x), \tag{3.26}$$

where

$$f_1(x) = (1-p)g(x-h(x))(K_{F,h}(x)+1)F(h(x))/g(x), \tag{3.27}$$

$$f_2(x) = (1-p)g(h(x))J_{F,h}(x)/g(x), \tag{3.28}$$

$$f_3(x) = \frac{(1-p)J_{F,h}(x) + (1-p^2)K_{F,h}(x) - (1-p)(K_{F,h}(x)+1)\overline{F}(h(x))}{g(x)}. \tag{3.29}$$

Considering (3.27) we apply (3.23) and (2.15) to find that

$$f_1(x) \to 1-p < 1.$$

Also, from (3.28) and (3.23), (2.16), we have

$$f_2(x) \to 0.$$

For (3.29) we apply (3.22) and (2.15) to obtain

$$f_3(x) \leq \varphi(b) \text{ for all } x \geq b. \tag{3.30}$$

Let

$$\delta(y) := \sup_{x \geq y}((f_1(x) + f_2(x)). \tag{3.31}$$

Since $\delta(y) \downarrow 1-p$ as $y \to \infty$ there exists $b > 0$ such that

$$\delta(b) < 1. \tag{3.32}$$

We can now state and prove

Theorem 3.4.1. *Let F be a subexponential distribution on $\mathbb{R}^+$, $h \in \hbar_F$ so that F is h-insensitive, and $g(x)$ satisfy conditions* (3.21),(3.22) *and* (3.23)*. Then, there exists $b > 0$, $0 < \delta(b) < 1$ and $\varphi(b) > 0$, such that for all $x \geq b$,*

$$\Delta_F(x) \leq \Delta_F^+(x) := Cg(x)$$

where

$$C = \max(\frac{\varphi(b)}{1-\delta(b)}, \varphi(b) + \delta(b)C[h(b), b]). \tag{3.33}$$

Proof. For $x \geq b$, $\Delta(x) \leq C(x)g(x) \leq C[b, x]g(x)$.

From (3.26) we know that

$$C(x) \leq (f_1(x) + f_2(x))C[h(x), x] + f_3(x).$$

Choose b as in (3.32) so that $\delta(b) < 1$ and $\varphi(b) < \infty$. Then

$$C[b, x] \leq \delta(b)C[h(b), x] + \varphi(b).$$

Now, $C[h(b), x] = \max(C[h(b), b], C[b, x])$.

If $C[h(b), x] = C[h(b), b]$ then

$$C[b, x] \leq \delta C[h(b), b] + \varphi(b).$$

If $C[h(b), x] = C[b, x]$ then

$$C[b, x] \leq \frac{\varphi(b)}{1-\delta(b)}.$$

This completes the proof. □

3.5 Applying the Result

Some numerical estimation of the tail distribution of the geometric sum must be done in order to evaluate $C[h(b), b]$. The greater the value of b, the more accurate the upper bound becomes. However, this comes at the greater computational cost of numerically evaluating the tail of the distribution. A compromise has to be struck between the tightness of the upper bound and the resources one is willing to invest in evaluating the tail.

A critical part of the procedure is the choice of the test function $g(x)$, which itself depends on the choice of $h(x)$. In [51] $g(x)$ was chosen as a function in closed form over the whole range of values of its argument. We observe, however, that in evaluating $C[h(b), b]$, we know the (numerically) exact value of $\Delta_F(x)$ in the range $h(b) \leq x \leq b$,

and we wish to use this information in our choice of $g(x)$ when appropriate. Our strategy in applying the result is therefore as follows.

1) Decide what resources are available for estimating the tail distribution to a suitable degree of accuracy. Given the available resources, define B as the maximum value for which we numerically evaluate $\Delta_F(B)$ and numerically evaluate $C[h(B), B]$.

2) Determine the class of functions that will do for $h(x)$.

3) Estimate $J(x, h(x))$ and $K(x, h(x))$ in the range $h(B) \leq x \leq B$.

4) Choose monotonically decreasing $g(x)$, which will depend on our particular choice for $h(x)$, and which may incorporate our numerical knowledge of $C(x)$, such that $g(x) = O(\max(J_{F,h}(x), K_{F,h}(x), \overline{F}(h(x))))$.

5) If $\sup_{x \geq B} \delta(x) < 1$, we take $b = B$, and find the corresponding value of φ. If $\sup_{x \geq B} \delta(x) \geq 1$ either the procedure has failed, or we must be prepared to use a larger value of B.

6) Calculate C.

Some comments on these steps will be useful.

In step 2 when choosing $h(x)$, there is a tension involved between the relative rates of decay of $J_{F,h}(x)$ and $K_{F,h}(x)$. The larger $h(x)$ is the smaller $J_{F,h}(x)$ becomes, but the larger $K_{F,h}(x)$ becomes, and vice versa. We can change the rates of decay of $J_{F,h}(x)$ and $K_{F,h}(x)$ by scaling $h(x)$ by some numerical factor without affecting the asymptotic decay rate of $g(x)$.

In the paper [51] Kalashnikov and Tsitsiashvili found a function h that ensured that, asymptotically, the decay rate of $J_{F,h}(x)$ and $K_{F,h}(x)$ was the same. We observe that there is more than one function h with this property. We exploit the simple device of finding one suitable function h and considering positive multiples of h and optimising over these choices.

In Step 4 we will generally want to choose $g(x)$ in order to cause the upper bound for the relative accuracy to decay to zero as fast as possible. We may also want to incorporate the information we have already calculated for $C[h(B), B]$. The fastest asymptotic decay rate for $g(x)$ is obtained by an optimal choice of $h(x)$. However, we can use the information we have gathered in calculating the numerically exact value of $\Delta_F(x)$ in the range $[h(b), b]$ by constructing a monotonically decreasing version of it, $\Delta_m(x) := \sup_{x \le y \le b^*} \Delta(y)$, for $x < b^*$, where $b^* < b$ is chosen to minimize the value of C. Thus, once we know the optimal asymptotic function $g(x)$, we instead use $g_1(x)$:

$$g_1(x) = \begin{cases} \Delta_m(x), & x < b^*, \\ Kg(x), & x \ge b^*, \end{cases}$$

for some constant K chosen to make $g_1(x)$ continuous at b^*. Continuity is not essential, but it ensures that the resulting function g is monotonic (decreasing), and ensures that the behaviour of $\delta(x)$ at the change over is not too erratic.

The effect of this is to reduce, for some values of x, the ratio of $C(x) = \Delta_F(x)/g(x)$ to 1. If b^* is correctly chosen this can cut out large values of $C(x)$, and hence reduce C.

Hence we have two parameters that we can adjust, the scale factor for $h(x)$, which alters the balance between $J_{F,h}(x)$ and $K_{F,h}(x)$ while keeping their asymptotic decay rates the same; and the value of b^*, which allow us, given our chosen value of B, to minimize C, and hence tighten the upper bound.

We will now show how to apply our result to (shifted) Pareto and Weibull distributions with various parameters. The values of $J_{F,h}(x)$ were estimated using the integrate function in R. The bounds we calculate are compared to values of the relative error that were calculated using a discretized Panjer algorithm with bandwidth of 0.005. Code for the R programs that performed the calculations can be found in Appendix C, together with brief descriptions.

3.5.1 Pareto Distribution

We will consider (shifted) Pareto Distributions of the following form:

$$F(x) = \begin{cases} 0, & x < 1, \\ 1 - x^{-\alpha}, & x \geq 1 \end{cases}$$

where $\alpha > 0$. We now follow the steps above. The choice of $h(x)$ is determined by the requirement that $K(x, h(x)) \to 0$. This occurs if and only if $h(x) = o(x)$. We then have

$$K(x, h(x)) = \left(1 - \frac{h(x)}{x}\right)^{-\alpha} - 1 \leq \frac{\alpha h(x) x^{\alpha}}{(x - h(x))^{\alpha+1}},$$
$$J(x, h(x)) \leq 2 \int_{h(x)}^{x/2} \frac{\overline{F}(x-y)}{\overline{F}(x)} F(dy) + \frac{\overline{F}^2(x/2)}{\overline{F}(x)} \leq 2 \left(\frac{2}{h(x)}\right)^{\alpha} + \left(\frac{4}{x}\right)^{\alpha}.$$

The simplest form of $h(x)$ is $h(x) = x^{\beta}$ for $0 < \beta < 1$. We can then choose

$$g(x) = x^{-\min(\alpha\beta, 1-\beta)},$$

which ensures that $g(x) = O(\max(J_{F,h}(x), K_{F,h}(x), \overline{F}(h(x))))$. If we want to make $g(x)$ decay as fast as possible the optimal choice for $h(x)$ will have $\beta = 1/(1+\alpha)$.

Example 1

We shall take $\alpha = 2.2$ and $p = 0.5$. We use a discretized Panjer recursion to estimate the tail of the distribution, and assume that our resources allow us to estimate this up to $B = 100$ using a bandwidth of 0.005.
If we follow the approach of [51] we take $h(x) = x^{1/3.2}$ and

$$g(x) = x^{-2.2/3.2} = x^{-0.6875}.$$

We find that $\delta(100) = 0.786$, and $C[5, 100] = 14.82$. This results in $C = 13.32$, giving

$$\Delta_F(x) \leq 13.52 x^{-0.6875}, \quad x > 100,$$

as the bound. (We only define the bound for $x > 100$ because our methodology assumes that we have calculated the exact relative error up to $B = 100$.)

Some improvement can be made by taking by scaling the function h to get $h(x) = 1.320x^{1/3.2}$, with a consequent value of $C = 12.8$.

A further improvement can be made by choosing the test function $g(x)$ to be equal to $\Delta_m(x)$, the monotonically decreasing version of the exact value of $\Delta(x)$, up to some value $b^* \leq B$, and then ensuring continuity at b^*. Each time a new value of b^* is proposed a fresh optimization has to be performed to find the optimal scaling factor for the function h. Optimizing over values of b^* and the scaling factor we obtain $h(x) = 1.14x^{1/3.2}$, and $b^* = 21.55$ as the optimal choice (for B=100), so that the function g becomes

$$g(x) = \begin{cases} \Delta_m(x), & x \leq 21.55, \\ 8.52x^{-0.6875}, & x > 21.55. \end{cases}$$

This results in the upper bound

$$\Delta_F(x) \leq 8.53x^{-0.6875}, \quad x > 100.$$

The logarithm of this upper bound has been graphed in Figure 3.4(a) along with the logarithm of the numerically exact result obtained from the Panjer recursion.

Example 2

Now we consider $\alpha = 2.2$ and $p = 0.2$. Once again we take $B = 100$. If we follow the methodology of [51] and take $h(x) = x^{1/3.2}$ and $g(x) = x^{-2.2/3.2}$, we find that $\min(n \in \mathbb{N} : \delta(n) < 1) = 1085$ which not only is greater than our chosen B, but far too large for the Panjer algorithm to deal with in any foreseeable future, or with any degree of accuracy once machine rounding errors are taken into account.

No improvement can be made just by scaling $h(x)$. However, if we apply the b^*-methodology the optimal choice (with B=100) is to take $h(x) = 1.05x^{1/3.2}$ and adjust $g(x)$ to coincide with $\Delta_m(x)$ for $x < b^* = 27.4$. This gives a bound, but we know that $\delta(x) \to 0.8$, and here, as we can see from Figure 3.5.1, $\delta(x)$ has only just decreased below 1. This implies that if it were 'economic' to choose a larger value for B then the bound could be improved. We can also see the erratic behaviour of δ as we through the 'changeover' in the definition of g.

The second part of the graph shows the value of $C(x)$, and we can see where the b^*-methodology has 'cut off' the largest values.

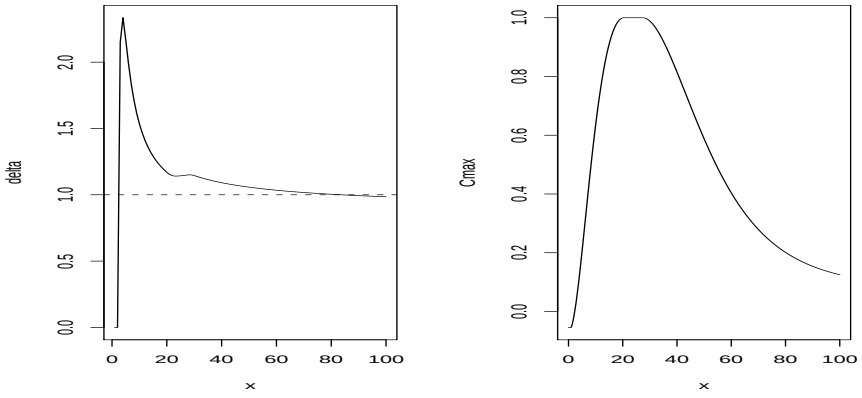


Figure 3.2: Example 2. Graphs of $\delta(x)$ and $C(x)$

We arrive at

$$g(x) = \begin{cases} \Delta_m(x), & x \leq 27.1, \\ 126x^{-0.6875}, & x > 27.1. \end{cases}$$

This gives the upper bound

$$\Delta_F(x) \leq 179.85x^{-0.687}, \quad x > 100,$$

which is shown in Figure 3.4(b).

Example 3

When the Pareto distribution in question is lighter tailed the asymptotic approximation becomes drastically less good for moderate values of the argument, and it requires more resources to compute numerically the tail distribution of the geometric sum for higher values of the argument.

As our example we take $\alpha = 5, p = 0.5$ and shall again perform this numerical exercise using a discretized Panjer algorithm, but still with $B = 100$ and a bandwidth of 0.002. Proceeding as in [51], we take $h(x) = x^{1/6}$ and $g(x) = x^{-5/6}$. We find that $\delta(100) = 0.867$, and $C = 1998$. This large value of C reflects the large values of $C(x)$ that

occur over a narrow range of values of x. Using the b^*-methodology we can cut these off. Optimizing we take $b^* = 30.58$ and $h(x) = 2.245x^{1/6}$ and

$$g(x) = \begin{cases} \Delta_m(x), & x \leq 30.58, \\ 42.26x^{-5/6}, & x > 27.1. \end{cases}$$

The effect of this change on $\delta(x)$ and $C(x)$ can be seen in Figure 3.3. It yields a dramatic improvement, giving

$$\Delta_F(x) \leq 42.36x^{-5/6}, \quad x > 100.$$

This bound is shown in Figure 3.4(c).

Example 4

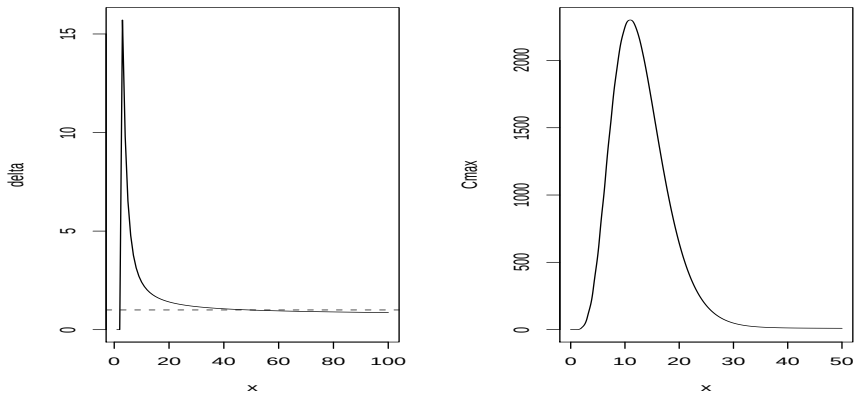
As a final example of a shifted Pareto, we consider the example from take Example 3.2.1, with $\alpha = 5, p = 0.2$. We again perform this numerical exercise using a discretized Panjer algorithm and a bandwidth of 0.002. For the first time we were unable to obtain results with $B = 100$ as, whatever choice we made for h and g we found that $\delta(100) > 1$.
However, with $B = 120$ we were able to find an optimal solution. Taking $h(x) = 1.693x^{1/6}$ and $b^* = 70.88$ we obtained the bound

$$\Delta_F(x) \leq 54059x^{-5/6}, \quad x > 120.$$

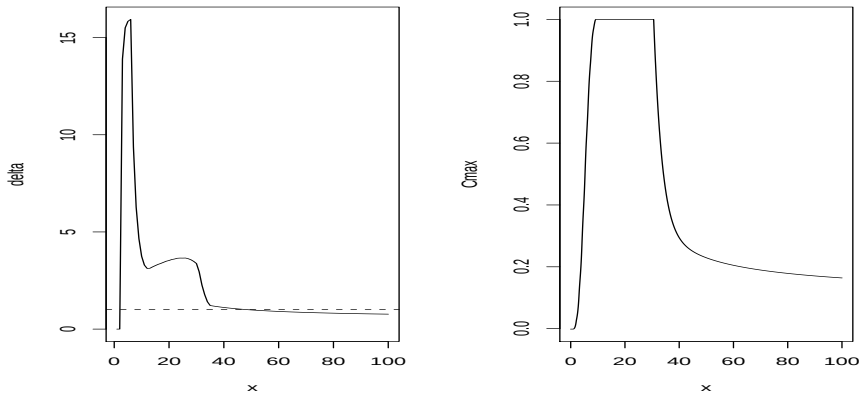
This bound is shown in Figure 3.4(d)
Increasing the value of B further to $B = 200$ we can improve this to

$$\Delta_F(x) \leq 6960x^{-5/6}, \quad x > 120,$$

but this is larger than the actual error by two orders of magnitude. So although even in this extreme case the method has found an upper bound it is not particularly good.



(a) Before adjustment



(b) After adjustment

Figure 3.3: Example 2. Graphs of $\delta(x)$ and $C(x)$ before and after

3.5.2 Weibull Distribution

We will consider a Weibull Distribution of the following form:

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-x^{\beta}}, & x > 0, \end{cases}$$

where $0 < \beta < 1$.

Straightforward calculations show that $h(x)$ must be chosen so that $h(x) = o(x^{1-\beta})$. Further calculations show that, if we take $(\log(x))^{1/\beta} = O(h(x))$, then $J_{F,h}(x) = O(K_{F,h}(x)$, and hence we may take $g(x) = K_{F,h}(x)$. Noting that

$$g(x) = O\left(\frac{(\log x)^2}{\sqrt{x}}\right),$$

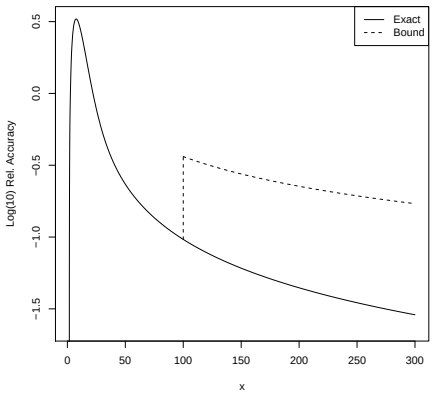
we also see that $\overline{F}(h(x)) = o(g(x))$.

Example 5

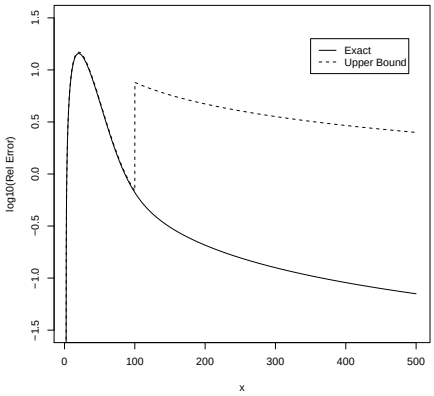
For this example we take $\beta = 0.5, p = 0.5$. The numerical calculations were done using a Panjer recursion with bandwidth 0.002, and B was taken to be $B = 100$. The optimal function for $h(x)$ is $h(x) = (\log(x))^2$. Applying the methodology in [51], no results can be obtained (for $B < 1660$). However, by taking $h(x) = 0.179(\log(x))^2$, we obtain

$$\Delta_F(x) \leq 2.952 K_{F,h}(x), \quad x \geq 50,$$

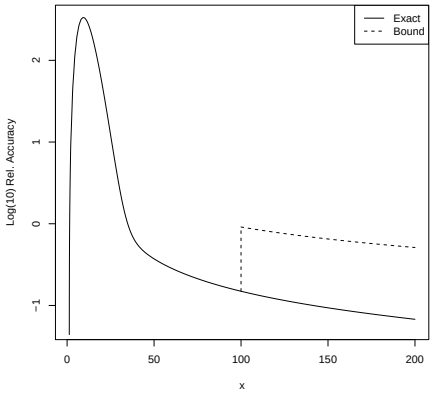
where $K_F(x) = \exp(\sqrt{x} - \sqrt{x - 0.179(\log(x))^2}) - 1$. This Weibull bound is shown in Figure 3.4(e). Because of the lack of a very sharp peak in $\Delta_F(x)$, no further improvement can be obtained using the b^* methodology.



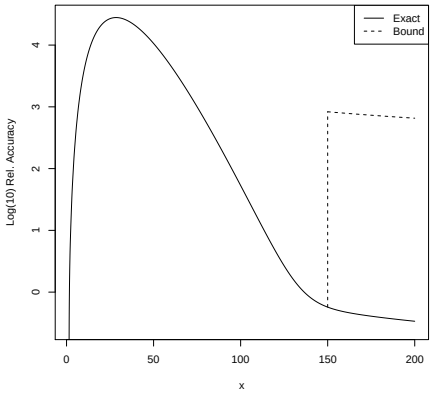
(a) Example 1.



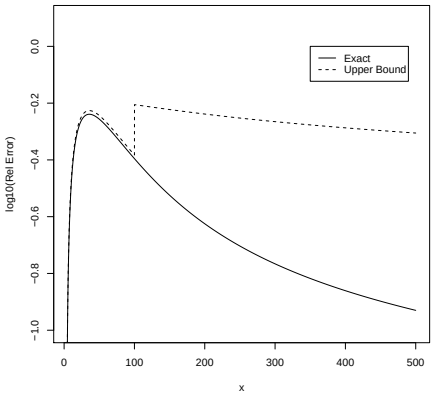
(b) Example 2.



(c) Example 3.



(d) Example 4.



(e) Example 5.

Figure 3.4: Plot of Upper bounds for Examples 1 - 5.

Chapter 4

Dependent Heavy-tailed Sums

Our analysis so far has concentrated on sums of independent heavy-tailed random variables. We now want to relax the condition of independence in order to deal with more general, and hopefully more realistic models. How easy it is to lose the independence between the summands was illustrated in the example in Section 1.3. Throughout this chapter we shall interpret heavy-tailed to mean subexponential; that is, we shall ignore the artificial examples that demonstrate that there is a difference between the heavy-tailed classes of distributions and concentrate on the 'natural' distributions. The key idea then from Chapter 2 is that such a sum will exceed a high threshold because of a single, very large jump, *the principle of the single big jump*. As should be clear, our interest is in the asymptotics of the tail of the distribution of the dependent heavy-tailed sum.

As we have said, for many practical purposes the independence assumption is too restrictive. Once we drop the requirement of independence, two questions naturally arise. First, *what kind of behaviours can occur as the dependence between the random variables strengthens?* And secondly, *how far beyond the independent case does the principle of the single big jump still hold?* These questions are of real interest, both from theoretical and practical viewpoints. We wish to consider the second question, and to establish conditions on the strength of the dependence which will preserve the results of the theory established for independent random variables outlined in Chapter

2; in particular, the principle of the single big jump.

Let X_1, X_2 be two subexponential random variables such that $\mathbf{P}(X_2 > x) \sim c\mathbf{P}(X_1 > x)$, where $0 \leq c \leq 1$; that is, we assume that X_1 has the heavier tail. Referring back to (1.15) we recall that we defined

$$\begin{aligned} P_1(x) =& \mathbf{P}(X_1 \wedge X_2 > x) \\ P_2(x) =& \mathbf{P}(X_1 \vee X_2 \leq x, X_1 + X_2 > x). \end{aligned}$$

By the decomposition in (1.15) we see that if $P_1(x)$ is negligible compared to $\mathbf{P}(X_1 > x)$, which in the independent case follows from the definition of subexponentiality, we have the Principle of the big jump. If in addition $P_2(x)$ is negligible compared to $\mathbf{P}(X_1 > x)$, as again is straightforward in the independent case, then we have the Principle of the single big jump. If the dependence is very strong, for instance if $X_1 = X_2$ a.s. (almost surely), then clearly the principle of the single big jump fails. In between the independent case and the comonotonic case a wide range of behaviours are possible. Intuition tells us that if the dependence is very light that the Principle of the single big jump should still hold. We wish to quantify in some way this notion of how light the dependence needs to be.

The remainder of this chapter is organized as follows. We first discuss some approaches that other authors have used in tackling these issues. We then introduce our approach, stating our assumptions. We state and prove a number of key theorems, concentrating on the case of non-negative sums. We then apply our results to a variety of different problems. In the next section we extend our results to the case of distributions supported on the whole real line. In the final section we discuss some heuristic methods which allow us to gain insight into the behaviour of dependent heavy-tailed sums without having to have recourse to the whole of our machinery.

4.1 Previous Approaches

A common approach to modelling dependence is the use of copulas, see, for example, [64]. A copula represents the dependence structure of a multivariate distribution by

transforming the distribution in such a way that the marginal distributions are all uniform on $(0, 1)$. Among the many authors who have used this methodology we mention [3, 4, 55, 12, 10]. Copulas have proved very popular in recent years and have lead to some important insights, but the methodology is not without its problems. Fitting a copula to data is a two stage problem. First one must fit the marginal distributions, which are then transformed to uniform distributions. It is not at all clear, at least to this author, why one should transform to a uniform marginal. More importantly one must then fit a particular copula structure to the uniform marginals. The possible space of copulas is infinite dimensional and very rich. In practice only a few classes of copulas are used, and they are used more for their mathematical convenience than for any overwhelming scientific reason. For a spirited if possibly overexuberant discussion of these and other issues related to copulas see [60].

Another approach has been to focus on notions of asymptotic tail dependence. For this we refer the reader to [61], although the analysis is restricted to distributions that lie in the maximum domain of attraction of the Gumbel extreme value distribution. Alternatively one may consider notions of positive or negative dependence. Here we mention [76, 53].

Alternatively some authors have considered bounds instead of asymptotics, for instance [29] Another common approach has been to concentrate on distributions with regularly varying marginals, for instance [46].

4.2 Insensitivity to Dependence and h-insensitivity

Before moving on to describe our main method, we consider a simple situation when the Principle of the single big jump holds, regardless of the strength of the dependence. Klüppleberg [52] observed that if X is a regularly varying random variable and Y is another random variable such that $\mathbf{P}(Y > o(x))) = o(\mathbf{P}(X > x))$, then $\mathbf{P}(X + Y > x) \sim \mathbf{P}(X > x)$ regardless of how strong any dependence between X and Y is; that is the tail of X is insensitive to the addition of Y. Casting this result in the light of

the notion of h-insensitivity we arrive almost straight away at the following result:

Proposition 4.2.1. *Let X be a long-tailed random variable and h be any function such that F is h-insensitive. Let Y be any random variable such that*

$$\mathbf{P}(Y > h(x)) = o(\mathbf{P}(X > x)) \tag{4.1}$$

then $\mathbf{P}(X + Y > x) \sim \mathbf{P}(X > x)$.

Proof. It is immediate that

$$\mathbf{P}(X + Y > x) \geq \mathbf{P}(X > x).$$

So consider

$$\begin{aligned}\mathbf{P}(X + Y > x) =& \mathbf{P}(X + Y > x, X \leq x - h(x)) + \mathbf{P}(X + Y > x, X > x - h(x)) \\ \leq& \mathbf{P}(Y > h(x)) + \mathbf{P}(X > x - h(x)) \\ \sim& \mathbf{P}(X > x),\end{aligned}$$

and hence the result follows. □

Clearly (4.1) is an increasing function property, so we can apply Proposition 2.3.2 and a simple induction argument to obtain

Proposition 4.2.2. *Let X be a random variable with subexponential distribution F and boundary class $\mathcal{H}$ generated by $H(x)$. Let $X_1, \ldots, X_n$ be r.v.s with distribution functions $F_1, \ldots, F_n$ such that, for all $i = 1, \ldots, n$ and all $c > 0$, $\overline{F}_i(cH(x)) = o(\overline{F}(x))$ then*

$$\mathbf{P}(X + X_1 + \ldots + X_n > x) \sim \mathbf{P}(X > x),$$

regardless of the dependence structure between the X's.

4.3 Conditionally Independent Random Variables

Our approach is to consider sums of random variables that are conditionally independent. That is, we condition on some information, represented by a sigma algebra, and

then, conditional on this information, the random variables in question are independent. Trivially this can always be done, because one can condition on what actually happens, at which point the random variables become degenerate on a single value and hence are independent. So we condition on the smallest sigma algebra that grants us independence. This allows great freedom, and in particular we need neither specify a particular subclass of subexponential distribution for which our results hold, nor assume the summands are identically distributed, nor specify any particular copula structure. However, the approach is not a universal panacea, and there are situations where the only sigma algebra that grants independence is the trivial sigma algebra.

This approach has the benefit that there are practical situations where a conditional independence structure arises naturally from the problem. As an example, consider a sequence of identical random variables $X_1, X_2, \ldots, X_n$, each with distribution function F_β depending on some parameter β that is itself drawn from a random variable drawn from an independent distribution. The X_i are independent once β is known: this is a typically Bayesian situation. It is natural to view the X_i as conditionally independent on the sigma algebra generated by β. We suppose the X_i to have subexponential (unconditional) distribution F and ask under what conditions the distribution of the sum follows the principle of the single big jump.

The approach developed from a paper of Foss, Konstantopoulos and Zachary [35] who studied time modulated random walks with heavy-tailed increments. In their proofs of two key theorems (Theorems 2.2 and 3.2) they used a coupling argument involving the sum of two conditionally independent random variables which entailed proving a lemma (Lemma A.2) which considered a particular simple case of conditional independence. The investigation in the present chapter considers this problem in much greater generality, whilst retaining the flavour of the simple situation in [35].

4.4 Conditions, Statements and Proofs

For the first time in this thesis it is necessary to be explicit about the probability model we are using so that we can define our σ-algebras; we work in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Initially we consider only non-negative random variables X_i, $i = 1, 2, \ldots$, each with distribution function (d.f.) F_i. We do not assume that each of these is subexponential, but we let F be a subexponential reference distribution on $\mathbb{R}^+$, and assume that each distribution is either proportionally tail equivalent to F or is asymptotically negligible compared to F. In this manner we avoid the complications that can arise owing to the complicated closure properties of subexponential distributions (complications that do not arise if we restrict ourselves either to the regularly varying distributions which we do not wish to do, or to non-negative subexponential distributions), see Section 2.2.2 for further details.

We let h be a function in $\hbar_F$ so that the reference distribution F is h-insensitive. It is important to appreciate that this does not imply that all of the distributions F_i are h-insensitive.

We make the following assumptions about the dependence structure of the X_i's:

(1) $X_1, X_2, \ldots$ are conditionally independent given $\mathcal{G} \subset \mathcal{F}$. That is, for any collection of indices $\{i_1, \ldots, i_r\}$, and any collection of sets $\{B_{i_1}, \ldots, B_{i_r}\}$, all belonging to $\mathcal{F}$, then $\mathbf{P}(X_{i_1} \in B_{i_1}, \ldots, X_{i_r} \in B_{i_r}|\mathcal{G})) = \mathbf{P}(X_{i_1} \in B_{i_1}|\mathcal{G})\mathbf{P}(X_{i_2} \in B_{i_2}|\mathcal{G}) \ldots \mathbf{P}(X_{i_r} \in B_{i_r}|\mathcal{G})$.

(2) For each $i \geq 1$, $\overline{F}_i(x) \sim c_i \overline{F}(x)$, with at least one $c_i > 0$.

(3) For each $i \geq 1$ there exists a non-decreasing functions $r(x)$ and an increasing collection of sets $B_i(x) \in \mathcal{G}$, with $B_i(x) \to \Omega$ as $x \to \infty$, such that

$$\mathbf{P}(X_i > x|\mathcal{G})\mathbf{1}(B_i(x)) \leq r(x)\overline{F}(x)\mathbf{1}(B_i(x)) \quad \text{almost surely.}$$

and, as $x \to \infty$, uniformly in i,

(4) $\mathbf{P}(\overline{B}_i(h(x))) = o(\overline{F}(x))$;

(5) $r(x)\overline{F}(h(x)) = o(1)$;

(6) $r(x)\int_{h(x)}^{x-h(x)} \overline{F}(x-y)F(dy) = o(\overline{F}(x))$.

Remark 4.4.1. When we are dealing with random sums of these random variables we need to have some uniformity condition on the distributions F_i. In this case we will amend condition (2) to condition

(2') For each $i \geq 1$, $\overline{F}_i(x) \sim c_i\overline{F}(x)$, with at least one $c_i > 0$ and so that for all $i \geq 1$ there exists $c > 0$ and $x_0 > 0$ such that $\overline{F}_i(x) \leq c\overline{F}(x)$ for all $x > x_0$.

Remark 4.4.2. Conditions (4-6) depend on being able to choose bounding functions $r(x)$ and bounding sets $B_i(x)$, which themselves depend on our choice of the function $h \in \hbar_F$. The choice of $h(x)$ is not unique, so the fact that one is unable to find appropriate bounding functions and sets for a particular function h does not imply that one cannot find them for some other choice of h. However, examination of conditions (3-5) reveals that they are all increasing function properties, in the sense of (2.20). Hence Proposition 2.3.2 implies that, as long as F possesses a boundary class, then instead of having to construct a function h for which the conditions do work, in practice we can work with any generator of the boundary class $\mathcal{H}_F$ and check the conditions with h replaced by all multiples of the generator. This makes the process of checking conditions (4-6) constructive and hence much easier.

Remark 4.4.3. In many cases the dependence between the $\{X_i\}$ enables us to choose a common $B(x) = B_i(x)$, for all i. However, we allow for situations where this is not the case. There is no need for a similar generality in choice of the function $r(x)$ because of the uniformity in i. The function $r(x)$ can be chosen so that it is only eventually monotone increasing, and in the case where we are only considering a finite collection of random variables $\{X_i\}$ it is sufficient to show that the chosen function is asymptotically equivalent to a monotone increasing function.

Remark 4.4.4. The need for and the meaning of the bounding functions $r(x)$ and the bounding sets $B_i(x)$ will become apparent when we give some examples. However, some preliminary comments may assist at this stage.

- In order to preserve the desired properties from the independent scheme, we need to ensure that the influence of the σ-algebra $\mathcal{G}$ that controls the dependence is not too strong. This we have done by introducing the bounding function $r(x)$ for the i^{th} random variable, which ensures that there are not events in $\mathcal{G}$ which totally predominate if a high level is exceeded. Although $r(x)$ may tend to infinity, it must not do so too quickly.

- Depending on the nature of the interaction of $\mathcal{G}$ with the random variables, there may be events in $\mathcal{G}$ that do overwhelmingly predominate when exceeding a high level; this is not a problem as long as these events are unlikely enough and their probability tends to zero as the level tends to infinity. Within the bounding sets $B_i(x)$ no events in $\mathcal{G}$ predominate, and we then require that the compliments $\overline{B}_i(x)$ decay quickly enough.

We have the following results.

Theorem 4.4.1. *Let X_i, $i = 1, 2, \ldots$ satisfy conditions (1-6) for some subexponential F concentrated on the positive half-line and for some $h(x) \in \hbar_F$. Then*

$$\mathbf{P}(X_1 + \cdots + X_n > x) \sim \sum_{i=1}^{n} \mathbf{P}(X_i > x) \sim \left(\sum_{i=1}^{n} c_i \right) \overline{F}(x).$$

Remark 4.4.5. Lemma A2 in [35] follows directly from this proposition.

In order to use dominated convergence to generalize Theorem 4.4.1 to random sums, we need the following extension of Kesten's Lemma 2.2.2.

Lemma 4.4.1. *With the conditions of (1-5), for any $\varepsilon > 0$ there exist $V(\varepsilon) > 0$ and $x_0 = x_0(\varepsilon)$ such that, for any $x > x_0$ and $n \geq 1$,*

$$\mathbf{P}(S_n > x) \leq V(\varepsilon)(1 + \varepsilon)^n \overline{F}(x).$$

Theorem 4.4.2. *With the conditions of (1), (2') and (3-6), let τ be an independent counting random variable such that $\mathbf{E}(e^{\gamma\tau}) < \infty$ for some $\gamma > 0$. Then*

$$\begin{aligned}\mathbf{P}(X_1 + \cdots + X_\tau > x) &\sim \mathbf{E}\left(\sum_{i=1}^{\tau} \mathbf{P}(X_i > x)\right) \\ &\sim \mathbf{E}\left(\sum_{i=1}^{\tau} c_i\right) \overline{F}(x).\end{aligned}$$

Clearly, checking that (5) and (6) hold is the most laborious part of guaranteeing the conditions for these propositions. Hence we propose a sufficient condition, analogous to the condition for subexponentiality in Proposition 2.2.5.

Proposition 4.4.1. *Let F be a subexponential distribution concentrated on the positive half-line, $h(x)$ be a function in $\hbar_F$, and $r(x)$ a non-decreasing function. Let $Q(x)$, the hazard function for F, be concave for $x \geq x_0$, for some $x_0 < \infty$. Let*

$$xr(x)\overline{F}(h(x)) \to 0 \quad \text{as } x \to \infty. \tag{4.2}$$

Then conditions (5) and (6) are satisfied.

Now we proceed with the proofs of our results.

Proof of Theorem 4.4.1. First consider $X_1 + X_2$. Assume, without loss of generality, that $c_1 > 0$. Let Y be a random variable, independent of X_1 and X_2 with distribution function F. We have the inequalities

$$\begin{aligned}\mathbf{P}(X_1 + X_2 > x) \leq &\mathbf{P}(X_1 > x - h(x)) + \mathbf{P}(X_2 > x - h(x)) \\ &+ \mathbf{P}(h(x) < X_1 \leq x - h(x), X_2 > x - X_1),\end{aligned}$$

and

$$\mathbf{P}(X_1 + X_2 > x) \geq \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x) - \mathbf{P}(X_1 > x, X_2 > x).$$

Now,

$$\begin{aligned}&\mathbf{P}(h(x) < X_1 \leq x - h(x), X_2 > x - X_1) \\ =&\mathbf{E}(\mathbf{P}(h(x) < X_1 \leq x - h(x), X_2 > x - X_1 | \mathcal{G}))\end{aligned}$$

$$
\begin{aligned}
=&\mathbf{E}\left(\int_{h(x)}^{x-h(x)}\mathbf{P}(X_1\in dy|\mathcal{G})\mathbf{P}(X_2>x-y|\mathcal{G})(\mathbf{1}(B_2(x-y))+\mathbf{1}(\overline{B}_2(x-y)))\right)\\
\leq& r(x)\mathbf{E}\left(\int_{h(x)}^{x-h(x)}\mathbf{P}(X_1\in dy|\mathcal{G})\mathbf{P}(Y>x-y)\right)+\mathbf{E}(\mathbf{1}(\overline{B}_2(h(x))))\\
=& r(x)\int_{h(x)}^{x-h(x)}\mathbf{P}(X_1\in dy)\overline{F}(x-y)+o(\overline{F}(x))\\
=& o(\overline{F}(x)).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\mathbf{P}(X_1>x,X_2>x)\\
=&\mathbf{E}(\mathbf{P}(X_1>x,X_2>x|\mathcal{G})(\mathbf{1}(B_2(x))+\mathbf{1}(\overline{B}_2(x))))\\
\leq&\mathbf{E}(\mathbf{P}(X_1>x|\mathcal{G})\mathbf{P}(X_2>x|\mathcal{G})\mathbf{1}(B_2(x)))+\mathbf{E}(\mathbf{1}(\overline{B}_2(x)))\\
\leq& r(x)\overline{F}(x)\mathbf{P}(X_1>x)+o(\overline{F}(x))\\
=& o(\overline{F}(x)).
\end{aligned}
$$

Hence, $\mathbf{P}(X_1+X_2>x)\sim\mathbf{P}(X_1>x)+\mathbf{P}(X_2>x)$. Since $c_1>0$, then $\mathbf{P}(X_1>x)+\mathbf{P}(X_2>x)\sim(c_1+c_2)\overline{F}(x)$.
Then, by induction, we have the desired result for the sum of n random variables. □

Proof of Lemma 4.4.1. Let Y be a random variable, independent of X_1 and X_2, and with subexponential reference distribution F on $\mathbb{R}^+$. For $x_0\geq 0$, which will be chosen later, and $k\geq 1$ put

$$\alpha_k=\alpha_k(x_0):=\sup_{x>x_0}\frac{\mathbf{P}(S_k>x)}{\overline{F}(x)}.$$

Also observe that

$$\sup_{0<x\leq x_0}\frac{\mathbf{P}(S_k>x)}{\overline{F}(x)}\leq\frac{1}{\overline{F}(x_0)}:=\alpha.$$

Take any $\varepsilon > 0$. Recall that for all $i > 0$, $\overline{F}_i(x) \leq c\overline{F}(x)$, for some $c > 0$ and for all $x > 0$. Then for any $n > 1$

$$\begin{aligned}\mathbf{P}(S_n > x) =& \mathbf{P}(S_{n-1} \leq h(x), X_n > x - S_{n-1}) \\ &+ \mathbf{P}(h(x) < S_{n-1} \leq x - h(x), X_n > x - S_{n-1}) \\ &+ \mathbf{P}(S_{n-1} > x - h(x), X_n > x - S_{n-1}) \\ \equiv& P_1(x) + P_2(x) + P_3(x).\end{aligned}$$

We bound

$$P_1(x) \leq \mathbf{P}(X_n > x - h(x)) \leq cL(x_0)\overline{F}(x)$$

and

$$P_3(x) \leq \mathbf{P}(S_{n-1} > x - h(x)) \leq \alpha_{n-1}L(x_0)\overline{F}(x)$$

for $x \geq x_0$, where $L(x) = \sup_{y \geq x} \frac{\overline{F}(y-h(y))}{\overline{F}(y)}$. For $P_2(x)$,

$$\begin{aligned}&P_2(x) = \mathbf{P}(h(x) < S_{n-1} \leq x - h(x), X_n > x - S_{n-1}) \\ &= \mathbf{E}\left(\int_{h(x)}^{x-h(x)} \mathbf{P}(S_{n-1} \in dy|\mathcal{G})\mathbf{P}(X_n > x - y|\mathcal{G})(\mathbf{1}(B_n(x-y)) + \mathbf{1}(\overline{B}_n(x-y)))\right) \\ &\leq \mathbf{E}\left(r(x)\int_{h(x)}^{x-h(x)} \mathbf{P}(S_{n-1} \in dy|\mathcal{G})\mathbf{P}(Y > x - y)\right) + \mathbf{P}(\overline{B}_n(h(x))) \\ &= r(x)\int_{h(x)}^{x-h(x)} \mathbf{P}(S_{n-1} \in dy)\mathbf{P}(Y > x - y) + \mathbf{P}(\overline{B}_n(h(x))) \\ &\leq r(x)\left(\int_{h(x)}^{x-h(x)} \mathbf{P}(Y \in dy)\mathbf{P}(S_{n-1} > x - y) + \mathbf{P}(S_{n-1} > h(x))\mathbf{P}(Y > x - h(x))\right) \\ &\quad + \mathbf{P}(\overline{B}_n(h(x))) \\ &\leq (\alpha_{n-1} + \alpha)r(x)\left(\int_{h(x)}^{x-h(x)} \mathbf{P}(Y \in dy)\mathbf{P}(Y > x - y) + \mathbf{P}(Y > h(x))\mathbf{P}(Y > x - h(x))\right) \\ &\quad + \mathbf{P}(\overline{B}_n(h(x))) \\ &= (\alpha_{n-1} + \alpha)\left(r(x)\int_{h(x)}^{x-h(x)} \frac{\overline{F}(x-y)}{\overline{F}(x)}F(dy) + r(x)\overline{F}(h(x))\overline{F}(x - h(x))\right) + \mathbf{P}(\overline{B}_n(h(x))).\end{aligned}$$

We now choose x_0 such that, for all $x \geq x_0$,

$$\begin{aligned}
\frac{\overline{F}(x-h(x))}{\overline{F}(x)} \leq L(x_0) \leq & 1+\frac{\varepsilon}{4};\\
r(x)\int_{h(x)}^{x-h(x)} \frac{\overline{F}(x-y)}{\overline{F}(x)}F(dy) \leq & \frac{\varepsilon}{4};\\
r(x)\overline{F}(h(x))L(x_0) \leq & \frac{\varepsilon}{4};\\
\frac{\mathbf{P}(\overline{B}_n(h(x)))}{\overline{F}(x)} \leq & 1
\end{aligned}$$

which can be done by virtue of the long-tailedness of F and conditions (3-6). We then have that

$$P_2(x) \leq \frac{\varepsilon}{2}(\alpha_{n-1}+\alpha)\overline{F}(x)+\overline{F}(x)$$

We therefore have

$$\begin{aligned}
\mathbf{P}(S_n > x) &\leq cL(x_0)\overline{F}(x)+\frac{\varepsilon}{2}(\alpha_{n-1}+\alpha)\overline{F}(x)+\overline{F}(x)+\alpha_{n-1}L(x_0)\overline{F}(x)\\
&\leq R\overline{F}(x)+(1+\frac{3}{4}\varepsilon)\alpha_{n-1}\overline{F}(x),
\end{aligned}$$

for some $0 < R < \infty$. Hence

$$\alpha_n \leq R+(1+\frac{3}{4}\varepsilon)\alpha_{n-1}.$$

Then, by induction we have

$$\begin{aligned}
\alpha_n &\leq \alpha_1(1+\frac{3}{4}\varepsilon)^{n-1}+R\sum_{r=0}^{n-2}(1+\frac{3}{4}\varepsilon)^r\\
&\leq Rn(1+\frac{3}{4}\varepsilon)^{n-1}\\
&\leq V(\varepsilon)(1+\varepsilon)^n,
\end{aligned}$$

for some constant $V(\varepsilon)$ depending on ε.
This completes the proof. □

Proof of Theorem 4.4.2. The proof follows directly from Theorem 4.4.1, Lemma 4.4.1 and the dominated convergence theorem. □

Proof of Proposition 4.4.1. Without loss of generality we may assume that $x_0 = 0$. Clearly (4.2) implies that condition (5) holds. Since Q is concave, the minimum of the sum $Q(x-y)+Q(y)$ on the interval $[h(x), x-h(x)]$ occurs at the endpoints of the interval. From Lemma 2.2.3, there exists a constant $C > 0$ such that

$$\begin{aligned}\int_{h(x)}^{x-h(x)} \overline{F}(x-y)F(dy) &\leq C\int_{h(x)}^{x-h(x)} \overline{F}(x-y)\overline{F}(y)dy \\ &= C\int_{h(x)}^{x-h(x)} \exp(-(Q(x-y)+Q(y)))dy \\ &\leq Cx\exp(-(Q(h(x))+Q(x-h(x)))) \\ &= Cx\overline{F}(h(x))\overline{F}(x-h(x)),\end{aligned}$$

and so

$$\begin{aligned}r(x)\int_{h(x)}^{x-h(x)} \frac{\overline{F}(x-y)}{\overline{F}(x)}F(dy) &\leq Cxr(x)\overline{F}(h(x))\frac{\overline{F}(x-h(x))}{\overline{F}(x)} \\ &= o(1).\end{aligned}$$

Therefore condition (6) also holds. □

4.5 Examples of Conditionally Independent Subexponential Random Variables

4.5.1 Example 1

This example is so straightforward that it can be tackled by the standard theory of independent subexponential random variables. Its purpose is didactic, and is designed to show why the bounding sets $B_i(x)$ are necessary.

Let ξ_i, $i = 1, 2, \ldots, n$, be i.i.d. with common distribution function $F_\xi \in R_{-\alpha}$. Let η be independent of the ξ_i and have distribution function $F_\eta \in R_{-\beta}$, where $\alpha \neq \beta$. Define $X_i = \xi_i + \eta$ for $i = 1, 2, \ldots, n$, and let the reference distribution be $\overline{F}(x) = x^{-(\alpha\wedge\beta)}$. Then, from Proposition 2.2.9 we have, for $i = 1, 2, \ldots, n$,

$$\mathbf{P}(X_i > x) \sim \overline{F}(x).$$

Conditional on the sigma algebra $\mathcal{G} = \sigma(\eta)$, the X_i are independent.

For our reference distribution, the boundary class is generated by $H(x) = x$. Now, the random variables $\frac{\mathbf{P}(X_i > x|\mathcal{G})}{\overline{F}(x)} \leq \frac{1}{\overline{F}(x)}$ are unbounded as $x \to \infty$. If we try to satisfy condition (3) without using the bounded sets $B(x)$, we need to ensure that, for all $x > 0$, almost surely,

$$\frac{\mathbf{P}(X_i > x|\mathcal{G})}{\overline{F}(x)} \leq r(x).$$

If we take $r(x) = \frac{1}{\overline{F}(x)}$, then condition (5) is not satisfied, since for any $c > 0$,

$$r(x)\overline{F}(H(x)) = c^{-(\alpha \wedge \beta)}.$$

Hence, we need to use bounded sets. Let $B(x) = \{\eta \leq x/2\}$. This satisfies condition (4), if and only if $\alpha < \beta$: for any $c > 0$,

$$\frac{\mathbf{P}(\overline{B}(H(x))}{\overline{F}(x)} = \frac{\mathbf{P}(\eta > cx/2)}{\overline{F}(x)} = (2/c)^{\beta} x^{(\alpha \wedge \beta) - \beta}.$$

Clearly, in the case $\alpha < \beta$, we may take $r(x)$ as a constant, $r(x) = 2^{\beta}$.

The condition $\alpha < \beta$ agrees with arguments on X_i taken from the standard theory of independent subexponential random variables.

4.5.2 Example 2

Let η be a random variable with uniform distribution in the interval$(1, 2)$. Conditional on $\mathcal{G} = \sigma(\eta)$ let X_i, $i = 1, 2, \ldots, n$, be i.i.d. with common distribution function

$$\overline{F}_{\xi|\eta}(x) = (1 + x)^{-\eta}, \quad x > 0.$$

Routine calculations show that

$$\mathbf{P}(X_i > x) \sim \frac{1}{x \log(1 + x)},$$

and we take $\overline{F}(x) = 1/(x \log(1 + x))$ to be our reference distribution. The boundary class is again generated by $H(x) = x$.

For all $x > 1$ we have, almost surely,

$$\begin{aligned}\frac{\mathbf{P}(X_i > x|\mathcal{G})}{\overline{F}(x)} &\leq \frac{\mathbf{P}(X_i > x|\eta = 1)}{\overline{F}(x)} \\ &= (1 + 1/x)\log(1 + x) \\ &\leq 2\log(1 + x),\end{aligned}$$

and we shall take $r(x) = 2\log(1 + x)$.

Routine calculations show that, for all $0 < c < 1/2$ condition (6) is satisfied, and also condition (5).

In this example there has been no need to define bounding sets, or equivalently we can take $B(x) = \Omega$ for all $x > 0$.

4.5.3 Example 3

In this example we again consider a Bayesian type situation. Let X_i, $i = 1, 2, \ldots, n$ be identically distributed, conditionally independent on parameter β, with conditional distribution F_β given by

$$\overline{F}_\beta(x) = \exp(-\gamma x^\beta), \quad \gamma > 0$$

where β is drawn from a uniform distribution on (a, b), $0 \leq a < 1$, $a < b$. The unconditional distribution of X_i, F_X, is then

$$\overline{F}_X(x) = \frac{1}{(b-a)\log x}\left(E_1(\gamma x^a) - E_1(\gamma x^b)\right),$$

where $E_1(x) = \int_x^\infty \frac{e^{-u}}{u} du$ is related to the exponential integral.

We now consider separately the two cases (i) $0 < a < 1$ and (ii) $a = 0$. We start first with case (i).

We find

$$\overline{F}_X(x) \sim \overline{F}(x) := \frac{\exp(-\gamma x^a)}{(b-a)\gamma x^a \log x}.$$

The boundary class $\mathcal{H}$ is the same as for the Weibull distribution,

$$\mathcal{H} = \{cx^{1-a},\ c > 0\}.$$

We shall take $B(x) = \Omega$ for all $x > 0$, and

$$r(x) = \gamma(b-a)x^a \log x.$$

Note that $Q(x) = -\log\left(\overline{F}(x)\right) = \log\left(\gamma(b-a)\right) + x^a + a\log x + \log\log x$ is convex, and that

$$xr(x)\overline{F}(H(x)) = \frac{x^{1-a^2}\log(x)\exp\left(-\gamma c^a x^{a(1-a)}\right)}{c^a \log(cx)} \to 0$$

as $x \to \infty$ for all $c > 0$. Hence, by Lemma 2.2.5, the conditions (3-6) are met, and the principle of the single big jump holds.

We now consider case (ii), with β distributed uniformly on the interval $(0, b)$. The reference distribution is now

$$\overline{F}(x) = \frac{E_1(1)}{b\log x} := \frac{k}{\log x},$$

where $E_1(1) \approx 0.21938$.

Since $\overline{F}$ is slowly varying it is x-insensitive. Therefore, in order to satisfy (5) we need to choose $r(x)$ such that $\lim_{x\to\infty} \frac{r(x)}{\log x} = 0$. For the bounded sets $B(x)$, the problems clearly occur near $\beta = 0$, so we may try sets of the form $B(x) = \{\beta \in (a(x), b)\}$, with $a(x) \to 0$ as $x \to \infty$.

To satisfy condition (3) we need, for each $x > 0$ and for $\beta \in (a(x), b)$,

$$\exp(-\gamma x^{a(x)}) < \exp(-\gamma x^{\beta}) \leq \frac{kr(x)}{\log(x)},$$

which cannot be true if both $a(x) \to 0$ and $\frac{r(x)}{\log x} \to 0$ as $x \to \infty$.

Hence the conditions (3-6) cannot be met.

The question now arises whether, in this case, the principle of the single big jump still holds. The answer is no. To see why, we again consider the representation (1.15), for the sum of two independent identically distributed subexponential random variables X_1, X_2. For simplicity we shall consider the case where $b = \gamma = 1$.

We have

$$P_1(x) = \int_0^1 d\beta \int_0^x \beta u^{\beta-1}\exp(-u^\beta)du \int_0^x \beta v^{\beta-1}\exp(-v^\beta)dv\mathbf{1}(u+v > x).$$

Making the substitution $u = xy, v = xz$, we have

$$\begin{aligned} P_1(x) &= \int_0^1 d\beta x^{2\beta} \int_0^1 \int_0^1 \beta^2 y^{\beta-1} z^{\beta-1} \exp(-x^\beta(y^\beta + z^\beta)) \mathbf{1}(y+z>1) dy dz \\ &\leq \int_0^1 d\beta x^{2\beta} \exp(-x^\beta) \int_0^1 \int_0^1 \beta^2 y^{\beta-1} z^{\beta-1} \mathbf{1}(y+z>1) dy dz \\ &= \int_0^1 d\beta x^{2\beta} \exp(-x^\beta) J(\beta), \end{aligned}$$

where $J(\beta) = \mathbf{P}(Y_1^{1/\beta} + Y_2^{1/\beta} > 1)$ and $Y_1, Y_2 \sim U(0,1)$ are i.i.d. As $\beta \to 0$, $J(\beta) \to 0$, so for any $\delta > 0$ there exists $\varepsilon > 0$ such that $J(t) \leq \delta$ for all $t \leq \varepsilon$. Hence,

$$\begin{aligned} P_1(x) &\leq \left(\int_0^\varepsilon + \int_\varepsilon^1 \right) d\beta x^{2\beta} \exp(-x^\beta) J(\beta) \\ &\leq \int_0^\varepsilon d\beta x^{2\beta} \exp(-x^\beta)\delta + o(\exp(-x^{\varepsilon/2})). \end{aligned}$$

But

$$\begin{aligned} \int_0^\varepsilon d\beta x^{2\beta} \exp(-x^\beta) &\leq \int_0^1 d\beta x^{2\beta} \exp(-x^\beta) \\ &= \frac{1}{\log(x)} \int_1^\infty t \exp(-t) dt \\ &= \frac{1}{2\log(x)}. \end{aligned}$$

Therefore

$$P_1(x) = o(\overline{F}(x)),$$

and the principle of the big jump holds. However,

$$P_2(x) = \int_0^1 \exp(-2x^\beta) d\beta = \frac{1}{\log(x)} \int_2^{2x} \frac{e^{-u}}{u} du$$

so that

$$\mathbf{P}(X_1 > x, X_2 > x) \sim \frac{E_1(1)}{E_1(2)} \mathbf{P}(X_1 > x).$$

Hence, the principle of the *single* big jump does not hold.

We note that this result could have been demonstrated by a straightforward application of Theorem 2.2 in [3], but we have preferred to show explicitly which parts of the representation (1.15) are not negligible.

4.5.4 Example 4

For $i = 1, 2, \ldots, n$ let $X_i = \xi_i \eta_1 \eta_2 \ldots \eta_i$, where the $\{\xi_i\}$ are i.i.d, and the $\{\eta_i\}$ are i.i.d. and independent of the $\{\xi_i\}$. Then conditional on the σ-algebra generated by $\{\eta_1, \ldots, \eta_n\}$ the $\{X_i\}$ are independent. Let the $\{\xi_i\}$ have common distribution function F in the intermediately regularly varying class, $F_\xi := F \in \text{IRV}$ and let the $\{\eta_i\}$ have common distribution function F that is rapidly varying, $F_{\eta_1} \in \mathcal{R}_{-\infty}$. This is related to the example given in [56]. In their example the $\{\xi_i\}$ were chosen to belong to the class $\mathcal{D} \cap \mathcal{L}$. We have chosen the slightly smaller class of intermediate regular variation because:

1. examples which lie in the $\mathcal{D} \cap \mathcal{L}$ class that do not lie in the IRV class are constructed in an artificial manner;

2. as we have observed in Proposition 2.3.4 the IRV class of functions (that are not long-tailed) has a common boundary class, and hence is suitable for general treatment under our methodology. In the unlikely event that F is long-tailed, although it does not have a boundary class, because it is x-insensitive the argument still goes through unchanged (but with $H(x)$ replaced by $h(x)$).

The boundary class for F is generated by $H(x) = x$.
By Lemma B.0.1 the class $\mathcal{R}_{-\infty}$ is closed under product convolution, hence for each $i = 1, 2, \ldots, n$ we have X_i is of the form $X_i = \xi_i \eta$ where the d.f. of η, $F_\eta \in \mathcal{R}_{-\infty}$. Then by Lemma B.0.2 each X_i has d.f. $\overline{F}_{X_i}(x) \asymp \overline{F}(x)$.
We now proceed to the construction of the bounding sets, $B(x)$. For condition (3) to hold we need to restrict the size of η. By Lemma B.0.3 we can choose $\varepsilon > 0$ such that $\overline{F_\eta}(x^{1-\varepsilon}) = o(\overline{F}(x))$. For such an ε we choose $B(x) = \{\eta \leq x^{1-\varepsilon}\}$. Then for any $H(x) = cx \in \mathcal{H}$, $0 < c < 1/2$, condition (4) requires

$$\mathbf{P}(\overline{B}(H(x)) = \overline{F}_\eta((cx)^{1-\varepsilon}) = o(\overline{F}_\eta(x^{1-\varepsilon}) = o(\overline{F}(x)),$$

as required.

Now consider condition (3):

$$\mathbf{P}(X_i > x|\mathcal{G})\mathbf{1}(B(x)) \leq r(x)\overline{F}(x)\mathbf{1}(B(x)).$$

This implies that the choice for $r(x)$ satisfies

$$\frac{\mathbf{P}(\xi_i > x/\eta|\eta \leq x^{1-\varepsilon})}{\overline{F}(x)} \leq \frac{\mathbf{P}(\xi_i > x^{\varepsilon})}{\overline{F}(x)} = \frac{\overline{F}(x^{\varepsilon})}{\overline{F}(x)} \leq r(x).$$

Taking $r(x) = \frac{\overline{F}(x^{\varepsilon})}{\overline{F}(x)}$, for any $H(x) = cx \in \mathcal{H}$,

$$r(x)\overline{F}(H(x)) = \frac{\overline{F}(x^{\varepsilon})}{\overline{F}(x)}\overline{F}(cx) = o(1),$$

and

$$r(x)\int_{cx}^{(1-c)x} \frac{\overline{F}(x-y)}{\overline{F}(x)}F(dy) \leq \frac{\overline{F}(x^{\varepsilon})\overline{F}(cx)}{\overline{F}^2(x)}\overline{F}(cx) = o(1).$$

Hence all the conditions (3-6) are met, and the principle of the single big jump holds; that is:

$$\mathbf{P}(X_1 + \cdots X_n > x) \sim \sum_{i=1}^{n} \overline{F}_{X_i}(x).$$

Remark 4.5.1. If, in addition, F is continuous, then Theorem 3.4 (ii) of [21] shows that the restriction $F_\eta \in \mathcal{R}_{-\infty}$ can be eased to $\overline{F}_\eta = o(\overline{F})$.

4.5.5 Example 5

In this example we consider random variables $X_1, \ldots, X_n$ with lognormal marginals. First we recall some facts about lognormally distributed random variables. A r.v. $X \sim LN(\mu, \sigma^2)$ if $X = e^Y$ and $Y \sim N(\mu, \sigma^2)$. The distribution function of X is $F_X(x) \sim \frac{\sigma}{\sqrt{2\pi}logx} exp\left(-\frac{1}{2\sigma^2}(logx - \mu)^2\right)$. If two r.v.s $X_1, \sim LN(\mu_1, \sigma_1^2)$, $X_2, \sim LN(\mu_2, \sigma_2^2)$ then X_1 has a heavier tail than X_2, in the sense that $\overline{F_{X_2}}(x) = o(\overline{F_{X_1}}(x))$, if and only if either $\sigma_2 < \sigma_1$ or both $\sigma_2 = \sigma_1$ and $\mu_2 < \mu_1$. The boundary class for $F_X(x)$ is generated by $H(x) = \frac{x}{logx}$. We observe that if X_1 has a heavier tail than X_2 then, for all $c > 0$, $\overline{F_{X_2}}(cH(x)) = o(\overline{F_{X_1}}(x))$, and by reference to Proposition 4.2.2 this suggests we need only consider the dependence structure as it relates to those X_i which have the heaviest tail.

So first let the r.v.s which have the heaviest distribution be $X_1 = e^{Y_1}, \ldots, X_m = e^{Y_m}$, each distributed with $X_i \sim LN(\mu, \sigma^2)$ with common distribution function F. We specify the dependence structure by assuming that $(Y_1, \ldots, Y_m) \sim MVN((\mu, \ldots, \mu), \Sigma)$ where Σ is of full rank. We perform a factor analysis and write each $Y_i = t_{i1}Z_1 + \cdots + t_{ik}Z_k + W_i$, where, for $1 \leq j \leq k$, the Z_j are i.i.d. standard normal and, independently , for $1 \leq i \leq m$, $W_i \sim N(\mu_i, \sigma_i^2)$ are independent normal r.v.s. Since we place no restriction on k and Σ is of full rank, this factor analysis can always be performed (non-uniquely) such that the W_i are non-degenerate, that is $\sigma_i > 0$ for all $i = 1, \ldots, m$.

We take $\mathcal{G}$ to be the sigma algebra generated by $Z_1, \ldots, Z_k$, and conditional on this the X_i are independent. Our reference distribution is F with $\mathbf{P}(X_i > x) \sim \overline{F}(x)$ for all $i = 1, \ldots, m$. Each X_i can be written as $X_i = \psi_i \xi_i$, where $\psi_i = e^{t_{i1}Z_1 + \cdots + t_{ik}Z_k} \sim LN(0, s_i^2)$ and $s_i^2 = t_{i1}^2 + \cdots + t_{ik}^2$ and $\xi_i = e^{W_i} \sim LN(\mu, \sigma_i^2)$. Then, for each i, $s_i^2 + \sigma_i^2 = \sigma^2$.

We choose bounding sets $B_i(x) = \{\psi_i \leq x^\delta\}$, where $1 > \delta > \max_i(\frac{s_i^2}{\sigma^2})$ ensures that $\mathbf{P}(\overline{B_i}(cH(x))) = o(\overline{F}(x))$ for all $c > 0$.

Then, given $B_i(x)$,

$$\mathbf{P}(X_i > x | B_i(x)) \leq \mathbf{P}(\xi_i > x^{1-\delta}) \leq \overline{F}(x^{1-\delta}),$$

for large enough x. So we may take

$$r(x) = \frac{\overline{F}(x^{1-\delta})}{\overline{F}(x)}$$

which is monotonically increasing for large enough x.

Since the lognormal reference distribution has a hazard function that is eventually concave, and, for all $c > 0$,

$$xr(x)\overline{F}(cH(x)) = x\frac{\overline{F}(x^{1-\delta})}{\overline{F}(x)}\overline{F}\left(\frac{cx}{logx}\right) \to 0$$

as $x \to \infty$, then we can apply Proposition 4.4.1 and conditions (5,6) hold. Hence $\mathbf{P}(X_1 + \cdots + X_m > x) \sim m\mathbf{P}(X_1 > x)$.

Now we apply Proposition 4.2.2, with $X_1 + \cdots + X_m$ in place of X and $X_{m+1}, \ldots, X_n$ as the lighter tailed r.v.s and conclude that

$$\mathbf{P}(X_1 + \cdots + X_n > x) \sim m\mathbf{P}(X_1 > x),$$

and the principle of the single big jump holds.

The example of lognormal random variables with Gaussian copula was studied in [10]. Although our results agree with the results in [10], there are some small differences in the assumptions. In [10] it is assumed that the whole dependence structure is a Gaussian copula, but in our setup the dependence of the lighter-tailed random variables is not specified. However, we feel that, in practice, this is an unimportant point. More importantly, we assume that the covariance matrix Σ for the heaviest random variables is of full rank. This is not assumed in [10]. The condition in [10] is that each pair $X_i = e^{Y_i}, X_j = e^{Y_j}$, where $i \neq j$ and $Var(Y_i) = Var(Y_j)$ has correlation between Y_i and Y_j of $\rho_{ij} < 1$. This does not imply that the covariance matrix is of full rank. Indeed, a simple example, for $i = 1, 2, 3$ is given by Y_1, Y_2 i.i.d. standard normals, and $Y_3 = (Y_1 + Y_2)/\sqrt{2}$. It is clear that our methodology cannot deal with this example, as only conditioning on the trivial sigma algebra can make these independent. In this sense the result in [10] is more general.

4.6 Real-valued Random Variables

We wish to extend our investigation beyond non-negative random variables to conditionally independent subexponential random variables taking real values. In order to deal with this situation we need to add another condition to those enumerated in Section 2 at (D1), (D2) and (D3). Again we let F be a reference subexponential distribution, and $h \in \hbar_F$.

(7) For each $i, j \geq 1$ we have that

$$\mathbf{P}\left(X_i > x + h(x), X_j \leq -h(x)\right) = o(\overline{F}(x)).$$

We then have the following extension of Theorem 4.4.1.

Theorem 4.6.1. *Let X_i, $i = 1, 2, \ldots$ be real-valued random variables satisfying conditions (1-7) for some subexponential F concentrated on the positive half-line and for some $h \in \hbar_F$. Then*

$$\mathbf{P}(X_1 + \cdots + X_n > x) \sim \sum_{i=1}^{n} \mathbf{P}(X_i > x) \sim \left(\sum_{i=1}^{n} c_i \right) \overline{F}(x).$$

Proof of Proposition 4.4.1. The proof follows the general outline of the proof of Theorem 4.4.1. The derivation of an upper bound for $\mathbf{P}(X_1 + X_2 > x)$ remains as in Theorem 4.4.1. For the lower bound we have:

$$\begin{aligned}
\mathbf{P}(X_1 + X_2 > x) \geq & \mathbf{P}(X_1 > x + h(x), X_2 > -h(x)) + \mathbf{P}(X_2 > x + h(x), X_1 > -h(x)) \\
& - \mathbf{P}(X_1 > x + h(x), X_2 > x + h(x)) \\
= & \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x) + o(\overline{F}(x)),
\end{aligned}$$

where we have used the long-tailedness of X_1 and X_2, condition (7) and $\mathbf{P}(X_1 > x + h(x), X_2 > x + h(x)) = o(\overline{F}(x))$ for the same reasons as in Theorem 4.4.1.
Hence $\mathbf{P}(X_1 + X_2 > x) \sim \mathbf{P}(X_1 > x) + \mathbf{P}(X_2 > x)$, and the rest of the proof follows by induction. □

We can see that the only reason for condition (7) is to deal with the lower bound, and hence no change is needed to show the two following generalisations of Lemma 4.4.1 and Theorem 4.4.2.

Lemma 4.6.1. *We let $\{X_i\}$ be as in Theorem 4.6.1 and satisfying conditions (1-7). Then, for any $\varepsilon > 0$, there exist $V(\varepsilon) > 0$ and $x_0 = x_0(\varepsilon)$ such that, for any $x > x_0$ and $n \geq 1$,*

$$\mathbf{P}(S_n > x) \leq V(\varepsilon)(1 + \varepsilon)^n \overline{F}(x).$$

Theorem 4.6.2. *If, in addition to the conditions of Lemma 4.6.1, τ is an independent counting random variable such that $\mathbf{E}(e^{\gamma\tau}) < \infty$ for some $\gamma > 0$, then*

$$\mathbf{P}(X_1 + \cdots + X_\tau > x) \sim \mathbf{E}\left(\sum_{i=1}^{\tau} \mathbf{P}(X_i > x) \right).$$

To show that condition (7) is both non-empty and necessary we construct an example where it fails to hold and the principle of the single big jump fails.

Example 6

Consider a collection of non-negative i.i.d. random variables $\{Z_i\}_{i\geq 0}$ such that each Z_i has the distribution of a generic independent non-negative random variable Z, and $\mathbf{P}(Z > x) = 1/x^\alpha$ for $x \geq 1$. Also consider a collection, independent of the Z_i, of non-negative i.i.d. random variables $\{Y_i\}_{i\geq 0}$ such that each Y_i has the distribution of a generic independent non-negative random variable Y, and $\mathbf{P}(Y > x) = 1/x^\beta$ for $x \geq 1$, where $\alpha > \beta > 1$. For $i \geq 1$ let $X_i = Z_i - Y_i Z_{i-1}$.

First we show that the X_i satisfy conditions (1-6).

For any $i \geq 1$ we have that $\mathbf{P}(X_i > x) \sim \overline{F}(x) := 1/x^\alpha$ for $x \geq 1$, and we recall that the boundary class for F is generated by $H(x) = x$. We take $\mathcal{G} = \sigma(\{Y_i Z_{i-1}\}_{i\geq 1})$, and then, conditional on $\mathcal{G}$, the X_i are independent. To show that condition (3) is met we need to examine the random variables $\mathbf{P}(X_i > x|\mathcal{G})$, so consider

$$\mathbf{P}(X_i > x|Y_{i+1}Z_i = w) \leq \mathbf{P}(Z_i > x|Y_{i+1}Z_i = w) = \mathbf{P}(Z > x|YZ = w).$$

We calculate that

$$\frac{\mathbf{P}(Z > x|YZ = w)}{\overline{F}(x)} = \begin{cases} x^\alpha \left(\frac{x^{-\alpha+\beta} - w^{-\alpha+\beta}}{1 - w^{-\alpha+\beta}}\right) & \text{for } 1 < x \leq w, \\ 0 & \text{for } x > c. \end{cases}$$

$$\leq x^\beta := r(x).$$

Clearly $r(x)\overline{F}(cx) = o(1)$ for all $0 < c < 1/2$.

Also, straightforward estimation shows that $\int_{cx}^{(1-c)x} \frac{\overline{F}(x-y)F(dy)}{\overline{F}(x)} = O(x^{-\alpha})$, and hence condition (6) is met. We take $B(x) = \Omega$ for all $x \geq 0$ so that there is nothing to show for condition (4).

We now consider condition (7). For any $i \geq 1$, and any $h(x) \in \hbar_F$,

$$\begin{aligned}
\mathbf{P}(X_i > x + h(x), X_{i+1} < -h(x)) =& \mathbf{E}\left(\mathbf{P}(Z_1 - W > x + h(x), Z_2 - Y_2 Z_1 < -h(x)|W)\right) \\
\geq& \mathbf{E}\left(\mathbf{P}(Z_1 > x + h(x) + W, Z_2 < Z_1 - h(x)|W)\right) \\
\geq& \mathbf{E}\left(\mathbf{P}(Z_1 > x + h(x) + W, Z_2 < x + W|W)\right) \\
=& \mathbf{E}\left(\mathbf{P}(Z_1 > x + h(x) + W)\mathbf{P}(Z_2 < x + W|W)\right).
\end{aligned}$$

Hence, by Fatou's lemma,

$$\begin{aligned}
&\liminf_{x\to\infty} \frac{\mathbf{P}(X_i > x + h(x), X_{i+1} < -h(x))}{\overline{F}(x)} \\
\geq &\liminf_{x\to\infty} \mathbf{E}\left(\frac{\mathbf{P}(Z_1 > x + h(x) + W|W)}{\mathbf{P}(Z_1 > x)}\mathbf{P}(Z_2 < x + W|W)\right) \\
\geq &\mathbf{E}\left(\liminf_{x\to\infty}\left(\frac{\mathbf{P}(Z_1 > x + h(x) + W|W)}{\mathbf{P}(Z_1 > x)}\mathbf{P}(Z_2 < x + W|W)\right)\right) \\
= &1,
\end{aligned}$$

and so condition (7) is not met.

Finally, we show that the conclusion of Theorem 4.6.1 fails in this example:

$$\begin{aligned}
\mathbf{P}(X_1 + \ldots + X_n > x) = &\mathbf{P}(Z_n + (1 - Y_n)Z_{n-1} + \ldots + (1 - Y_2)Z_1 - Y_1 Z_0 > x) \\
\leq &\mathbf{P}(Z > x),
\end{aligned}$$

and hence the Principle of the single big jump does not hold.

4.7 Heuristic Methods

It is possible in some specific cases to gain insight into whether or not the Principle of the single big jump holds without having to have recourse to the whole of our machinery developed in Section 4.4. As an example consider a simplified version of Example 5 from Section 4.5.

Example 4.7.1.

Again, we consider random variables $X_1, \ldots, X_n$ with lognormal marginals. For each $i = 1, \ldots, n$, let $X_i = e^{W+Y_i}$, where $Y_1, \ldots, Y_n$ are i.i.d. normal with $Y_i \sim N(\mu, \sigma_1^2)$ and $W \sim N(0, \sigma_2^2)$, independent of the Y_i. Also, let $\xi_i = e^{Y_i} \sim LN(\mu, \sigma_1^2)$ and $\eta = e^W$. To see why we expect the sum $X_1 + \cdots + X_n$ to satisfy the Principle of the single big jump we consider the most likely way that a large level of the sum is exceeded.

We see that $X_1 + \cdots + X_n = W(Z_1 + \cdots + Z_n)$. Now $\Xi := \xi_1 + \cdots + \xi_n$ is the sum of i.i.d. subexponential random variables, so from the definition of subexponentiality

we know that, asymptotically, the sum exceeds level x because one of ξ_i exceeds level x, and the others are all 'small'.

Now consider the product $\eta\Xi$ and consider how this becomes large. We know that $\Xi = e^{Y_1+\cdots+Y_n} := e^Z$, so that $Z \sim N(n\mu, n\sigma_1^2)$. So the product becomes large if the sum $W + Z$ becomes large. Now, we know from the properties of the normal distribution, or more generally from the properties of superexponential distributions, that a sum becomes large because, asymptotically, each of the summands contributes a specific proportion of the sum. That is, both W and Z are large (in a ratio that is determined in this case by the ratio of their variances).

Putting this together, the sum $X_1 + \cdots + X_n$ becomes large because both W and Z become large. But $\Xi = e^Z$ becomes large because precisely one of its summands ξ_i is large. Hence $X_1 + \cdots + X_n = \eta\xi_1 + \eta\xi_n$ becomes large because precisely one of its summands becomes large, and the Principle of the single big jump holds.

The idea of appealing to the most likely way that the sum becomes large is a natural way to think about subexponential sums. It can also free us from some of the restrictions from conditions (1) and (2) in Section 4.4.

Example 4.7.2.

Consider once again Example 5 from Section 4.5. As we remarked at the end of this example, if $X_i = e^{Y_i}$, for $i = 1, 2, 3$, have lognormal marginals, Y_1, Y_2 are i.i.d. standard normals, and $Y_3 = (Y_1 + Y_2)/\sqrt{2}$ then it is clear that our main methodology cannot deal with this example, as only conditioning on the trivial sigma algebra can make these independent. At first sight it seems strange that the Principle of the single big jump holds. However, to see why this is the case we again consider how the sum is most likely to become large. First we note that the marginal distributions of X_1, X_2, X_3 are identical. Then, because X_1 and X_2 are independent only one of these becomes large if their sum becomes large. If we consider how X_3 becomes large, then this is because both Y_1 and Y_2 are asymptotically the same size, and so are X_1 and X_2, but this implies that their sum cannot be large. Hence we can only have one of

X_1, X_2, X_3 large at any one time. This is clearly a heuristic argument, and provides insight rather than proof. A more rigorous approach to these types of argument needs to be developed.

This method, once it has been made rigorous should also help to deal with some situations that meet condition (1) but not condition (2); that is, although the marginal distributions may be subexponential, conditional on the sigma algebra the conditional distributions may not be. An example would be:

Example 4.7.3. Let $X_1, \ldots, X_n$ be non-negative random variables such that for $i = 1, \ldots, n$, $X_i = \eta\xi_i$ where $\eta, \xi, \xi_1, \ldots, \xi_n$ are i.i.d. random variables with common distribution $\overline{G}(x) = e^{-x}$, the standard exponential distribution.
First we argue that the X_i have subexponential distribution. We consider the typical way that $\eta\xi$ becomes large by considering how $\log\eta + \log\xi$ becomes large. The random variables $\log\eta$ and $\log\xi$ are i.i.d. with common distribution that is doubly exponential (there are two senses of doubly exponential distributions: the Laplace distribution, and the (left) tail of the Gumbel distribution; we mean the latter). This doubly exponential distribution is much lighter than the normal, and in our terminology is superexponential, and the sum of two such independent summands asymptotically concentrates around the diagonal, so considering the event $\log\eta + \log\xi > \log x$, we know that $\log\eta \approx \log\xi \approx \log(x)/2$, which suggests that $\mathbf{P}(\eta\xi > x) \approx e^{-2\sqrt{x}}$. In fact, this argument has only given the correct logarithmic asymptotics. To get the exact asymptotics we follow [75] and evaluate the following expression, making the change of variable $u = x/y + y - 2\sqrt{x}$ in (4.3):

$$\begin{aligned}\mathbf{P}(\eta\xi > x) &= \left(\int_0^{\sqrt{x}} + \int_{\sqrt{x}}^{\infty}\right) e^{-x/y} e^{-y} dy \\ &= \int_0^{\infty} e^{-(u+2\sqrt{x})} \left(-du + d\sqrt{u^2 + 4u\sqrt{x}}\right)/2 \\ &\quad + \int_0^{\infty} e^{-(u+2\sqrt{x})} \left(du + d\sqrt{u^2 + 4u\sqrt{x}}\right)/2 \qquad (4.3)\end{aligned}$$

So,

$$\begin{aligned}
\mathbf{P}(\eta\xi > x) =& \int_0^\infty e^{-(u+2\sqrt{x})} d\sqrt{u^2 + 4\sqrt{x}u} \\
=& \frac{1}{2} x^{-1/4} e^{-2\sqrt{x}} \int_0^\infty u^{1/2} e^{-s} \frac{\sqrt{4u\sqrt{x}}}{\sqrt{u^2 + 4u\sqrt{x}}} du \\
&+ x^{1/4} e^{-2\sqrt{x}} \int_0^\infty u^{-1/2} e^{-s} \frac{\sqrt{4u\sqrt{x}}}{\sqrt{u^2 + 4u\sqrt{x}}} du \\
\sim& \frac{\sqrt{\pi}}{4} x^{-1/4} e^{-2\sqrt{x}} + \sqrt{\pi} x^{1/4} e^{-2\sqrt{x}} \qquad (4.4) \\
\sim& x^{1/4} e^{-2\sqrt{x}},
\end{aligned}$$

where in (4.4) we have used the dominated convergence theorem as $x \to \infty$.
Now conditional on the σ-algebra generated by η the conditional distribution is exponential. The sum can be written as

$$X_1 + \cdots + X_n = \eta(\xi_1 + \cdots + \xi_n).$$

Both η and $\xi_1 + \cdots + \xi_n$ have exponential-like tails, and there is no concentration effect , either around the diagonal, or on the co-ordinate axes for $(\xi_1 + \cdots + \xi_n)$. Also $(\xi_1 + \cdots + \xi_n)$ has a heavier tail than η, so there will be a concentration effect around a large value of $\xi_1 + \cdots + \xi_n$, which in turn implies that all possible combinations of values of $\xi_1, \ldots, \xi_n$ contribute significantly and hence heuristics suggest that the Principle of the single big jump fails in this case.

Appendix A

Notations and Conventions

In this appendix we list the notations and conventions we have used.

Sets of Numbers $\mathbb{R}$ is the set of real numbers. $\mathbb{R}^+$ is the set of non-negative numbers. $\mathbb{Z}$ are the integers. $\mathbb{Z}^+ = \{0, 1, \ldots\}$.

Intervals $[a, b]$ represents a closed interval and (a, b) an open interval. Corresponding notation represents half-open intervals.

Indicator function $\mathbf{1}(A)$ stands for the indicator function of event A, so that $\mathbf{1}(A) = 1$ if A holds, and $\mathbf{1}(A) = 0$ otherwise.

Probability and Expectation With respect to some underlying probability space $\mathbf{P}(A)$ represents the probability of event A. $\mathbf{E}(X)$ represents the expectation of a random variable X and $\mathbf{E}(X; A)$ stands for the expectation of X over the event A, that is $\mathbf{E}(X; A) = \mathbf{E}(X\mathbf{1}(A))$.

Distributions If F stands for the distribution of a random variable X then, for any event A, $F(A) = \mathbf{P}(X \in A)$. The distribution function is $F(x) = \mathbf{P}(X \leq x)$ and the tail distribution is $\overline{F}(x) = \mathbf{P}(X > x)$.

Convolutions If F and G are the distributions, respectively, of two independent random variables X and Y then the convolution $F * G$ is the distribution of $X + Y$. F^{*n} then stands for the n-fold convolution of F with itself.

Limit convention Throughout, unless stated otherwise, all limit relations are for $x \to \infty$.

Landau notation Let $a(x)$ and $b(x)$ be two positive functions such that

$$0 \leq l_1 = \liminf_{x\to\infty} \frac{a(x)}{b(x)} = \limsup_{x\to\infty} \frac{a(x)}{b(x)} \leq l_2 \leq \infty.$$

We write $a(x) = O(b(x))$ if $l_2 < \infty$ and $a(x) = o(b(x))$ if $l_2 = 0$. We say that $a(x)$ and $b(x)$ are weakly equivalent, written $a(x) \asymp b(x)$, if both $l_1 > 0$ and $l_2 < \infty$, and that $a(x)$ and $b(x)$ are (strongly) equivalent, written $a(x) \sim b(x)$, if $l_1 = l_2 = 1$.

Maximum and minimum $X \vee Y$ stands for $\max(X, Y)$, and $X \wedge Y$ stands for $\min(X, Y)$.

Definition The symbol $:=$ means the quantity on the left is defined to be equal to the quantity on the right.

Equality in distribution $X \stackrel{d}{=} Y$ means that the two random variables are equal in distribution.

Appendix B

Regular Variation and Related Distributions

We collect here some definitions of well-known classes of functions and distributions. We also state three lemmas that are used in Example 4 in Section 4.5.4.

Slow Variation A positive function l is slowly varying if, for all $\lambda > 0$, $l(\lambda x) \sim l(x)$.

Regular Variation A distribution function F belongs to the class of regularly varying distributions of degree α, $\mathcal{R}_{-\alpha}$, if $\overline{F}(x) \sim l(x)x^{-\alpha}$ for some slowly varying function l.

Extended Regular Variation A distribution function F belongs to the class of extended regular varying distributions, ERV, if $\liminf_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} \geq \lambda^{-c}$ for some $c \geq 0$ and all $\lambda \geq 1$.

Intermediate Regular Variation A distribution function F belongs to the class of intermediately regular varying distributions, IRV, also called consistent variation by some authors, if $\lim_{\lambda\downarrow 1} \liminf_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = 1$.

Dominated Variation A distribution function F belongs to the class of dominatedly varying distributions, $\mathcal{D}$, if $\liminf_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} \geq 0$ for some $\lambda > 1$.

Rapid Variation A distribution function F belongs to the class of rapidly varying distributions, $\mathcal{R}_{-\infty}$, if $\lim_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = 0$ for all $\lambda \geq 1$.

Relationship between the different classes We write $\mathcal{S}$ for the class of subexponential distributions, and $\mathcal{L}$ for the class of long-tailed distributions. Then we have

the proper inclusions (see [32]):

$$\mathcal{R}_{-\alpha} \subset \text{ERV} \subset \text{IRV} \subset \mathcal{D} \cap \mathcal{L} \subset \mathcal{S} \subset \mathcal{L}.$$

The following three lemmas are due to [21, 77, 56] and are used in the development of the example in 4.5.4.

Lemma B.0.1. *The class $\mathcal{R}_{-\infty}$ is closed under product convolution.*

Lemma B.0.2. *Let X and Y be two independent positive r.v.s, and let the distribution function of X, $F_X \in \mathcal{D} \cap \mathcal{L}$, and that of Y, $F_Y \in \mathcal{R}_{-\infty}$. Let the distribution function of XY be F_{XY}. Then $\overline{F}_{XY}(x) \asymp \overline{F}_X(x)$.*

Lemma B.0.3. *If $F \in \mathcal{D}$ and $F_\eta \in \mathcal{R}_{-\infty}$ then there exists $\varepsilon > 0$ such that $\overline{F_\eta}(x^{1-\varepsilon}) = o(\overline{F}(x))$.*

Appendix C

R Code

This appendix contains the R code that was used to generate the numerical work in Chapter 3.1.

Details of the main functions that are used by the practitioner are given. Invisible output is not described in detail. The functions are listed in the order they are designed to be used in. In particular, any analysis must start with `geo.dist`.

Functions that are designed only to be called by other functions are not given detailed descriptions, but code is still given.

geo.dist

Description

This function generates the exact tail distribution of the geometric sum using a discretized Panjer algorithm. It compares the numerical value to the asymptotic expression and calculates the relative error of the approximation.

It then plots the logarithm of the relative error.

Usage

```
geo.dist(p,xu=300,bdw=0.05,type,alpha=0,lam=1,beta=0)
```

Arguments

`p`	The geometric parameter
`xu`	The upper value for which the exact distribution is calculated

`bdw`	The bandwidth for the discretization of the Panjer algorithm
`type`	character string giving the type of distribution
	pareto1 is ordinary pareto
	pareto2 is a Pareto shifted to start at 1
	weibull is usual Weibull
`alpha`	Pareto shape parameter. Must be > 0
`lam`	Location parameter for pareto1 or weibull type distribution
`beta`	Weibull shape parameter

Values

Invisible output for passing to other functions. A graph of logarithm of the relative error.

Code

```
geo.dist = function(p,xu=300,bdw=0.05,type,alpha=0,lam=1,beta=0){
x=seq(0,xu,by=bdw)
len=length(x)
g=numeric(len)
q=1-p
if (type=="pareto1") {
Fbar=function(x) pareto.Fbar1(x,alpha,lam)
f=function(x) pareto.f1(x,alpha,lam)
}
if (type=="pareto2") {
Fbar=function(x) pareto.Fbar2(x,alpha)
f=function(x) pareto.f2(x,alpha)
}
if (type=="weibull") {
Fbar=function(x) weibull.Fbar(x,beta,lam)
f=function(x) weibull.f(x,beta,lam)
```

```
}
Fbar.v=Fbar(x)
dF=-c(0,diff(Fbar.v))
denom=1-q*dF[1]
g[1]=p*dF[1]/denom
for (i in 2:len) g[i]=(p*dF[i]+q*sum(dF[2:i]*g[(i-1):1]))/denom

Gbar=1-cumsum(g)
exact=(p*Gbar-Fbar.v)/Fbar.v
ex.mon=exact
imax=which(exact==max(exact[is.finite(exact)]))[1]
ex.mon[1:imax]=exact[imax]
par(mfrow=c(1,1))
ymi=log10(0.8*min(ex.mon))
yma=log10(1.2*max(ex.mon))
plot(x[exact>0],log10(exact[exact>0]),type="l",ylim=c(ymi,yma),
xlab="x",ylab="Log(10) Rel. Acc.",main="Rel. Acc. of Asymptotic")
par(mfrow=c(1,1))
invisible(list(x=x,tail=Gbar,Fbar=Fbar,f=f,exact=exact,ex.mon=ex.mon,
p=p,xu=xu,bdw=bdw,alpha=alpha,beta=beta,lam=lam,type=type))
}
```

KT.table

Description

This function selects the functions $h(x)$ and $g(x)$, then calculates the functions $J_{F,h}(x)$ and $K_{F,h}(x)$. It then finds $\delta(x)$, $C(x)$, and C. It plots two graphs, one for $\delta(x)$ and the other for $C(x)$.

Usage

```
KT.table(ex.d,xl,xu,by,tol,hC=1,gb=0)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`xl`	The lower value for which $\delta(x)$, $C(x)$ are calculated.
	Must be > 0. Recommended is $xl = 1$
`xu`	The upper value for which $\delta(x)$, $C(x)$ are calculated.
	Must be greater than texttttol
`by`	The step size from `xl` to `xu`
	at which the functions are evaluated
`tol`	The tolerance parameter B
`hC`	The scale factor multiplying the function $h(x)$.
	Default value is 1
`gb`	The b^* constant for adjusting the function $g(x)$.
	Default value is 0

Values

Two graphs of $\delta(x)$ and $C(x)$.

`delta`	The x value at which $\delta(x)$ first becomes < 1
	and the corresponding value of $\delta(x)$
`tol.v`	The values of B, $\delta(B)$ and C
`Cmax`	The x value at which $C(x)$ reaches its maximum
	and the corresponding value of $C(x)$
`gconst`	The constant k which ensures continuity of $g(x)$
	at the join, and the value of kC

Code

```
KT.table=function(ex.d,xl,xu,by,tol,hC=1,gb=0){
if (ex.d$type=="pareto1") fns=fp(ex.d,hC,gb)
if (ex.d$type=="pareto2") fns=fp(ex.d,hC,gb)
if (ex.d$type=="weibull") fns=fw(ex.d,hC,gb)
```

```
p=ex.d$p;q=1-p
Fbar=ex.d$Fbar
f=ex.d$f
h=fns$h
if (gb==0) {g=fns$g.plain;k=1}
if (gb>0) {g=fns$g.adj;k=fns$k}
J=function(x) Jf(x=x,hx=h(x),Fb=Fbar,f=f)
K=function(x) Kf(x=x,hx=h(x),Fb=Fbar)
x=seq(xl,xu,by=by);len=length(x)
itol=max(which(x<=tol))
Fhx=Fbar(h(x))
gx=g(x);gh=g(h(x));gxh=g(x-h(x))
Jx=J(x);Kx=K(x)
delta=(q*gxh*(Kx+1)*(1-Fhx)+q*gh*Jx)/gx
phi=(q*Jx+(1-p^2)*Kx-q*(Kx+1)*Fhx)/gx
test=KT.max(ex.d,g,tol)
C.max=test$C.max
par(mfrow=c(1,2))
plot(x,delta,type="l")
abline(1,0,lty=2)
plot(test$x.test,test$C.test,type="l",xlab="x",ylab="Cmax")
par(mfrow=c(1,1))
if (delta[itol]>=1) stop("\n Tolerance too low \n")
if (delta[len]<1) idelta=max(which(delta>=1))+1
else {stop("\n Not converged \n")}
C=pmax(phi/(1-delta),C.max*delta+phi)
delta.con=cbind(x[idelta],delta[idelta],C[idelta])
;colnames(delta.con)=c("x","delta","C")
tol.v=cbind(x[itol],delta[itol],C[itol])
```

```
colnames(tol.v)=c("x","delta","C")
Cmax=cbind(test$x.max,C.max);colnames(Cmax)=c("x","Cmax")
gconst=cbind(k,k*C[itol]);colnames(gconst)=c("k","kC")
return(list(delta=delta.con,tol.v=tol.v,Cmax=Cmax,gconst=gconst))
}
```

opt1

Description

This function optimizes the scale-factor hC that multiplies $h(x)$ to obtain the smallest possible value of kC. It may be used with or without the b^*-methodology. If it returns

Usage

```
opt1(ex.d,tol,hCl,hCu,gb=0)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`hCl`	The lower value of the range in which the routine searches for the optimum
`hCu`	The upper value of the range in which the routine searches for the optimum
`gb`	The b^* constant for adjusting the function $g(x)$. Default value is 0

Values

`minimum`	The value of hC that minimizes the value of kC in the specified range
`objective`	The minimized value of kC

Code

```
opt1=function(ex.d,tol,hCl,hCu,gb=0){
hC.f=function(hC) C.calc(ex.d=ex.d,x=tol,hC=hC,gb=gb)
optimize(hC.f,c(hCl,hCu))
}
```

opt2

Description

This function performs a two dimensional optimization for the scale-factor hC that multiplies $h(x)$ and the b^* constant for adjusting the function $g(x)$ to obtain the smallest possible value of kC. It must start at a point where $\delta(B) < 1$

Usage

```
opt2(ex.d,tol,hCstart=1,gbstart))
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`hCstart`	The starting value for hC
`gbstart`	The starting value for gb

Values

`par`	[1] the value of hC and [2] the value of b^* that minimize the value of kC
`value`	The minimized value of kC

Code

```
opt2 = function(ex.d,tol,hCstart=1,gbstart){
hC.f2=function(par) {if (par[1]<=0|par[2]<0) Inf
else {C.calc(ex.d=ex.d,x=tol,hC=par[1],gb=par[2])}}
optim(c(hCstart,gbstart),hC.f2)
}
```

opt.s

Description

This function performs a crude grid optimization to find a starting point for the use of `opt2`. As a crude optimization it is less efficient than `opt2`

Usage

```
opt.s(ex.d,tol,bl,bu,by, hCl,hCu)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`bl`	The lower value of the range for b^* in which the routine searches for the optimum
`bu`	The upper value of the range for b^* in which the routine searches for the optimum
`by`	The step size from `bl` to `bu`
`hCl`	The lower value of the range for hC in which the routine searches for the optimum
`hCu`	The upper value of the range for hC in which the routine searches for the optimum

Values

`par`	The value of hC and the value of b^* that minimize the value of kC
`value`	The minimized value of kC

Code

```
opt.s=function(ex.d,tol,bl,bu,by, hCl,hCu) {
bs=seq(bl,bu,by=by)
kB=numeric(length(bs))
hC=numeric(length(bs))
```

```
for (i in 1:length(bs))
{
op=opt1(ex.d,tol=tol,hCl=hCl,hCu=hCu,gb=bs[i])
hC[i]=op$minimum
kB[i]=op$objective
}
imin=which(kB==min(kB))
par=c(hC[imin],bs[imin])
value=kB[imin]
return(list(par=par,value=value))
}
```

KT.opt1

Description

This function takes the output from `opt1` and runs `KT.table`

Usage

```
KT.opt1(ex.d,tol,best)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`best`	Variable holding output from `opt1`

Values

As for `KT.table`

Code

```
KT.opt1=function(ex.d,tol,best) KT.table(ex.d=ex.d,xl=1,xu=tol,
by=1,tol=tol,hC=best,gb=0)
```

KT.opt2

Description

This function takes the output from either `opt2` or `opt.s` and runs `KT.table`

Usage

```
KT.opt2(ex.d,tol,best)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`best`	Variable holding output from `opt2` or `opt.s`

Values As for `KT.table`

Code

```
KT.opt2=function(ex.d,tol,best) KT.table(ex.d=ex.d,xl=1,xu=tol,
by=1,tol=tol,hC=best$par[1],gb=best$par[2])
```

best.gr

Description

This function takes the output from either `opt2` or `opt.s` and plots the upper bound against the numerical value of the relative error. It also summarises all information

Usage

```
best.gr(ex.d,tol,best)
```

Arguments

`ex.d`	The variable containing the output from `geo.dist`
`tol`	The tolerance parameter B
`best`	Variable holding output from `opt2` or `opt.s`

Values

`alpha`	The Pareto α parameter
`beta`	The Weibull β parameter

`hC`	The value of hC
`gamma`	The power for the Pareto $h(x)$
`k`	The value of k
`C`	The value of C
`kC`	The value of kC
`theta`	The power for the Pareto $g(x)$
`gb`	The value of b^* for the b^*-adjustment

Code

```
type=ex.d$type
x=ex.d$x
xu=ex.d$xu
exact=ex.d$exact
ex.mon=ex.d$ex.mon
alpha=ex.d$alpha
beta=ex.d$beta
Fbar=ex.d$Fbar
hC=best$par[1]
gb=best$par[2]
kC=best$value
gamma=1/(alpha+1)
theta=gamma-1
ib=min(which(x>=gb))
itol=min(which(x>=tol))
if (type=="pareto1") {
ub=function(x) kC*x^theta
k=ex.mon[ib]/gb^theta
}
if (type=="pareto2")  {
ub=function(x) kC*x^theta
```

```
k=ex.mon[ib]/gb^theta
}
if (type=="weibull") {
g=function(x) Fbar(x-hC*log(x)^(1/beta))/Fbar(x)-1
ub=function(x) kC*g(x)
k=ex.mon[ib]/g(gb)
}
C=kC/k
ymi=log10(0.8*min(ex.mon))
yma=log10(max(1.2*ub(tol),ex.mon))
par(mfrow=c(1,1))
plot(x[exact>0],log10(exact[exact>0]),type="l",xlab="x",
ylab="Log(10) Rel. Accuracy",ylim=c(ymi,yma))
xg=seq(tol,xu,length=1000)
lines(xg,log10(ub(xg)),lty=2)
lines(c(tol,tol),c(log10(ex.mon[itol]),log10(ub(tol))),lty=2)
legend(x="topright",c("Exact","Bound"),lty=c(1,2))
if (type=="pareto1")
list(alpha=alpha,hC=hC,gamma=gamma,k=k,C=C,kC=kC,theta=theta,gb=gb)
if (type=="pareto2")
list(alpha=alpha,hC=hC,gamma=gamma,k=k,C=C,kC=kC,theta=theta,gb=gb)
if (type=="weibull")
list(beta=beta,hC=hC,gamma=gamma,k=k,C=C,kC=kC,theta=theta,gb=gb)
}
```

fp

Description

This function returns $h(x)$ and either the plain or adjusted $g(x)$ for Pareto type distributions

Code

```
fp = function(ex.d,hC,gb) {
alpha=ex.d$alpha
gamma=1/(alpha+1)
theta=gamma-1
h=function(x) pmin(x/2,hC*x^gamma)
k=1
ib=min(which(ex.d$x>=gb))
if (gb>0) k=ex.d$ex.mon[ib]/gb^theta
g.plain=function(x) x^theta
g.adj= function(x){
g=k*x^theta
for (i in 1:length(x)) if (x[i] < gb)
g[i]=ex.d$ex.mon[max(which(ex.d$x<=x[i]))]
return(g)
}
invisible(list(h=h,g.plain=g.plain,g.adj=g.adj,k=k))
}
```

fw

Description

This function returns $h(x)$ and either the plain or adjusted $g(x)$ for Weibull type distributions

Code

```
fw = function(ex.d,hC,gb) {
Fbar=ex.d$Fbar
beta=ex.d$beta
gamma=1/beta
```

```
h=function(x) pmin(x/2,hC*log(x)^gamma)
K=function(x) Kf(x=x,hx=h(x),Fb=Fbar)
k=1
ib=min(which(ex.d$x>=gb))
if (gb>0) k=ex.d$ex.mon[ib]/K(gb)
g.plain=function(x) K(x)
g.adj= function(x){
g=k*K(x)
for (i in 1:length(x)) if (x[i] < gb)
g[i]=ex.d$ex.mon[max(which(ex.d$x<=x[i]))]
return(g)
}
invisible(list(h=h,g.plain=g.plain,g.adj=g.adj,k=k))
}
```

KT.max

Description

This function calculates $C(x)$ and finds its maximum value in $[0, B]$

Code

```
KT.max=function(ex.d,g,tol){
exact=ex.d$exact[ex.d$x<=tol]
x.test=ex.d$x[ex.d$x<=tol]
C.test=exact/g(x.test)
C.max=max(C.test[is.finite(C.test)])
im=which(C.test==C.max)
x.max=x.test[im]
list(x.test=x.test,C.test=C.test,x.max=x.max,C.max=C.max)
}
```

KT.max

Description

This function does the same as `KT.max` but only returns the maximum value of $C(x)$ in $[0, B]$

Code

```
quick.max=function(ex.d,g,tol){
test= (ex.d$x == floor(ex.d$x)) & (ex.d$x<=tol)
C.test=ex.d$exact[test]/g(ex.d$x[test])
C.max=max(C.test[is.finite(C.test)])
return(C.max)
}
```

Jf

Description

This function calculates $J_{F,h}(x)$

Code

```
Jf=function(x,hx,Fb,f){
len=length(x)
p.J=numeric(len)
for (i in 1:len){
integrand=function(w) Fb(x[i]-w)/Fb(x[i])*f(w)
p.J[i]=integrate(integrand,hx[i],x[i]-hx[i])$value}
return(p.J)
}
```

Kf

Description

This function calculates $K_{F,h}(x)$

Code

```
Kf=function(x,hx,Fb) Fb(x-hx)/Fb(x)-1
```

C.calc

Description

This function calculates an approximate value of kC for use in optimization loops.

Code

```
C.calc=function(ex.d,x,hC,gb){
if (ex.d$type=="pareto1") fns=fp(ex.d,hC,gb)
if (ex.d$type=="weibull") fns=fw(ex.d,hC,gb)
if (ex.d$type=="pareto2") fns=fp(ex.d,hC,gb)
p=ex.d$p;q=1-p
Fbar=ex.d$Fbar
f=ex.d$f
h=fns$h
if (gb==0) {g=fns$g.plain;k=1}
if (gb>0) {g=fns$g.adj;k=fns$k}
J=function(x) Jf(x=x,hx=h(x),Fb=Fbar,f=f)
K=function(x) Kf(x=x,hx=h(x),Fb=Fbar)
C.max=quick.max(ex.d,g,x)
delta=(q*g(x-h(x))*(K(x)+1)*(1-Fbar(h(x)))+q*g(h(x))*J(x))/g(x)
phi=(q*J(x)+(1-p^2)*K(x)-q*(K(x)+1)*Fbar(h(x)))/g(x)
if (delta < 1) C=max(phi/(1-delta),C.max*delta+phi)
else C=Inf
return(k*C)
}
```

Distribution functions and their densities

```
pareto.Fbar1=function(x,alpha,lam) (lam/(lam+x))^alpha
pareto.f1=function(x,alpha,lam) alpha/lam*(lam/(lam+x))^(alpha+1)

pareto.Fbar2=function(x,alpha) pmin(1,x^(-alpha))
pareto.f2=function(x,alpha) alpha*x^(-alpha-1)*(x>1)

weibull.Fbar=function(x,beta,lam) exp(-lam*x^beta)
weibull.f=function(x,beta,lam) lam*beta*x^(beta-1)*exp(-lam*x^beta)
```

Bibliography

[1] J. Abate, G.Choudhury and W. Whitt, Waiting time tail probabilities in queues with long-tail service-time distributions, Queueing Systems 16 (1994) 311-338.

[2] J. Abate and W. Whitt, Numerical inversions of Laplace transforms of probability distributions, ORSA J. Comp. 7, (1995) 36-43.

[3] H. Albrecher, S. Asmussen and D. Kortschak, Tail asymptotics for the sum of two heavy-tailed dependent risks, Extremes 9, (2006) 107-130.

[4] S. Alink, M. Löwe and M. Wüthrich, Diversification of aggregate dependent risks, Insurance: Math. Econom. 35, (2004) 77-95.

[5] S. Asmussen, *Applied Probability and Queues*, (Springer, New York, 2003).

[6] S. Asmussen and K. Binswanger, Simulation of ruinprobabilities for subexponential claims, ASTIN Bulletin 27, (1997) 297-318.

[7] S. Asmussen, K. Binswanger and B. Hojgaard, Rare Event Simulation for Heavy-tailed Distributions, Bernouilli 6, (2000) 303-322

[8] S. Asmussen and P. Glynn, *Stochastic Simulation* (Springer, New York, 2007).

[9] S. Asmussen and C. Klüppleberg, Large deviation results for subexponential tails, with applications to insurance risk, Stoch. Processes and their Applications 64 (1996) 103-125.

[10] S. Asmussen and L. Rojas-Nanayapa, Asymptotics of sums of lognormal random variables with Gaussian copula, Statistics and probability Letters 78, (2008) 2709-2714.

[11] K.B. Athreya and P.E. Ney, *Branching Processes*, (Springer-Verlag, Berlin, 1972).

[12] R. Baird, C. Lef'evre and S. Loisel, Impact of correlation crises in risk theory: asymptotics of finite-time ruin probabilities when some independence and stationarity assumptions are relaxed, Insurance: Math. Econom. 43, (2008) 412-421.

[13] A. Baltrunas, E. Omey, S. Van Gluck, Hazard rates and subexponential distributions, Publ. Inst. Math. (N.S.) 80 (94), (2006) 29-46.

[14] A. Bassamboo, S. Juneja and A. Zeevi, On the efficiency loss of state-independent importance sampling in the presence of heavy-tails, Op. Reseach Letters 34, (2006) 521-531.

[15] J. Bertoin and R. Doney , On the local behaviour of ladder height distributions, J. Appl. Probab. 31, (1994) 816-821.

[16] N. Bingham, C. Goldie, J. Teugels, *Regular Variation*, (Cambridge University Press, Cambridge, 1987).

[17] J. Blanchet and P. Glynn, Efficient rare-event simulation for the maximum of heavy-tailed random walks, Ann. Appl. Probab. 18, (2008), 1351-1378

[18] A. A. Borovkov, *Asymptotic Methods in Queuing Theory*, (Wiley, Chichester, 1984).

[19] J. Chover, P. Ney and S. Wainger, Functions of probability measures, J. Anal. Math. 26, (1973) 255-302.

[20] A. A. Borovkov, *Stochastic Processes in Queuing Theory*, (Springer, New York, 1976).

[21] D. Cline and G. Samorodnitsky, Subexponentiality of the product of independent random variables, Stoch. Processes and their Applications 46, (1994) 75-98.

[22] V.P. Chistyakov, A theorem on sums of independent positive random variables and its application to ranching processes, Theor. Probab. Appl. 9 (1964) 640-648.

[23] C. Cocozza-Thivent and V. Kalashnikov, The failure rate in reliability; approximation and bounds, J. Appl. Math. Stochastic Anal. 9, (1996) 497-530.

[24] H. Cramér, *On the Mathematical Theory of Risk*, Skandia Jubilee Vol. (Stockholm, 1930).

[25] D. Daley, E. Omey and R. Vesilo, The tail behaviour of a random sum of subexponential random variables and vectors, Extremes 10, (2007) 21 - 39.

[26] L. de Haan, On regular variation and its application to weak convergence of sample extremes, CWI Tract 32 (1970).

[27] F. de Vylder, A practical solution to the problem of ultimate ruin probability, Scan. Act. J. 2, (1978) 135-159.

[28] D. Denisov, S. Foss and D. Korshunov, Tail asymptotics for the supremum of a random walk when the mean is not finite, Queueing Systems 46 (2004) 1533.

[29] M. Denuit, C.Genest and 'E. Marceau, Stochastic bounds on sums of dependent risks, Insurance: Math. Econom. 25, (1999) 85-104.

[30] D. Dickson and H. Waters, The Distribution of the Time to Ruin in the Classical Risk Model, ASTIN Bulletin 32 (2002) 299-313.

[31] P. Embrechts and C. Goldie, On closure and factorization properties of subexponential distributions, J. Austral. Math. Soc. A 29, (1980) 243-256.

[32] P. Embrechts, C. Klüppleberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, (Springer, New York, 1997).

[33] P. Embrechts and N. Veraverbeke, Estimates for the probability of ruin with special emphasis on the possibility of large claims, Insurance: Math. Econom. 1, (1982) 55-72.

[34] W. Feller, *An Introduction to Probability Theory and Its Applications II*, (Wiley, New York, 1971).

[35] S. Foss, T. Konstantopoulos and S. Zachary, The principle of a single big jump: discrete and continuous time modulated random walks with heavy-tailed increments, J. Theor. Prob. 20, (2007) 581-612.

[36] S. Foss and D. Korshunov, Lower limits and equivalences for convolution tails, Ann. Probab. 35, (2007) 366383.

[37] S. Foss, D. Korshunov and S. Zachary, *An Introduction to Heavy-tailed and Subexponential Distributions*, Oberwolfach Preprints 2009.

[38] S. Foss and A. Puhalskii, Random walks and Lévy processes conditioned not to overshoot To appear.

[39] S. Foss and A. Richards, On Sums of Conditionally Independent Subexponential Random Variables, To appear in Math. Op. Research

[40] S. Foss and S. Zachary, The maximum on a random time interval of a random walk with long-tailed increments and negative drift, Annals of Applied Probability 13 (2003) 37-53.

[41] Foss, S.and Zachary Z., On the exact distributional asymptotics for the supremum of a random walk with increments in a class of light-tailed distributions, Siberian Math. J., Volume 47, Number 6, November 2006 , pp. 1034-1041(8).

[42] H. Gerber and F. Dufresne, Three methods to calculate the probability of ruin, ASTIN Bulletin 19, (1989) 71-90.

[43] B. Gnedenko and I. Ushakov, *Probabilistic Reliability Engineering*, (J. Wiley and Sons, New York, 2003).

[44] C. Goldie, Subexponential distributions and dominated-varying tails, J. Appl. Probab. 15, (1978) 440-442.

[45] M. Goovaerts and F. De Vylder, A stable recursive algorithm for evaluation of ultimate ruin probabilities, ASTIN Bulletin 14, (1984) 53-59.

[46] M. Goovaerts, R. Kaas, Q.Tang, and R.Vernic, The tail probability of discounted sums of Pareto-like losses in insurance, *Proceedings of 8th Int. Congress on Insurance: Mathematics and Economics*, (Rome, 2004)

[47] J. Grandell, *Aspects of Risk Theory*, (Springer, New York, 1991).

[48] S. Juneja and P. Shahabuddin, Simulating heavy tailed processes using delayed hazard rate twisting, ACM TOMACS 12, (2002) 94-118.

[49] V. Kalashnikov, *Geometric Sums: Bounds for Rare Events with Applications*, (Kluwer Academic Publishers, Dordrecht, 1997).

[50] V. Kalashnikov and G. Tsitsiashvili, Tails of waiting times and their bounds, Queuing Systems 32 (1999) 257-283.

[51] V. Kalashnikov and G. Tsitsiashvili, Asymptotically Correct Bounds of Geometric Convolutions with Subexponential Components, J. Math. Sciences, 106 (2001) 2806-2819.

[52] C. Klüppleberg, Estimates of ruin probabilities by means of hazard rates, Insurance: Math. Econom. 8, (1989) 279-285.

[53] B. Ko and Q.Tang, Sums of dependent nonnegative random variables with subexponential tails, J. Appl. Prob. 45 (2008) 85-94

[54] D. Korshunov, The critical case of the Cramer-Lundberg theorem on the asymptotic tail behavior of the maximum of a negative drift random walk, Siberian Math. J. 46,(2005) ,10771081.

[55] D. Kortschak and H. Albrecher, Asymptotic results for the sum of dependent non-identically distributed random variables, Methodology and Computing in Applied Probability, to appear.

[56] R. Laeven, M. Goovaerts and T. Hoedemakers, Some asymptotic results for sums of dependent random variables, with actuarial applications, Insurance: Math. Econom. 37 (2005) 154-172.

[57] J.R. Leslie, On the non-closure under convolution of the subexponential family, J. Appl. Probab. 26, (1989) 58-66.

[58] F. Lundberg, Some supplementary researches on the collective risk theory, Skandinavisk Aktuarietidskrift 15, (1932) 137158.

[59] B. Mandelbrot, The Variation of Certain Speculative Prices, Journal of Business 36, (1963) 394-419.

[60] T. Mikosch, Copulas: Tales and Facts, Extremes 9, (2006) 3-20.

[61] A. Mitra and S. Resnick, Aggregation of risks and asymptotic independence, Eprint arXiv 0807.4200 (2008).

[62] M. Miyazawa and V. Schmidt, On ladder height distributions of general risk processes, Ann. Appl. Prob. 3, (1993) 763-776.

[63] E. Murphee, Some new results on the subexponential class, J. Appl. Prob. 26, (1989) 892-897.

[64] R. Nelsen *An Introduction to Copulas*, (Springer, New York, 2006).

[65] J. Neslehova, P. Embrechts and V. Chavez-Demoulin, Infinite mean models and the LDA for operational risk, Journal of Operational Risk 1, (2006) 3-25.

[66] A. Pakes, On the tails of waiting-time distributions, J. Appl. Probab. 12, (1975) 555-564.

[67] H. Panjer, Recursive evaluation of a family of compound distributions, ASTIN Bulletin 12 (1), (1981) 2226.

[68] Direct calculation of ruin probabilities, J. Risk Ins. 53, (1986) 521-529.

[69] E. Pitman, Subexponential distribution functions, J. Austral. Math. Soc. A 29, (1980) 337-347.

[70] Poisson, S. Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédé des règles gènèrales du calcul des probabilités, Paris: Bachelier, 1837.

[71] S. Resnick, *Extreme Values, Regular Variation and point Processes*, (Springer, New York, 1987).

[72] A. Richards, On Upper Bounds for the Tail Distribution of Geometric Sums of Subexponential Random Variables, Queuing Systems 62 (1999) 229-242.

[73] T. Rolski, H. Schmidli, V. Schmidt and J. Teugels, *Stochastic Processes for Insurance and Finance*, (Wiley, Chichester, 1999).

[74] G. Samorodnitsky, *Long Range Dependence*, (now publishers Inc., 2007).

[75] Q. Tang, From light tails to heavy tails through multiplier, Extremes 11, (2008) 379-391.

[76] Q. Tang, Insensitivity to negative dependence of the asymptotic behaviour of precise large deviations, Electronic Journal of Probability 11 (2006) 107-120.

[77] Q. Tang and G. Tsitsiashvili, Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks, Stoch. Processes and their Applications 108 (2003) 299-325.

[78] J. Teugels, The class of subexponential distributions, Ann. Prob. 3, (1975) 1001-1011.

[79] J. Teugels and N. Veraverbeke, Cramér-type estimates for the probability of ruin, CORE Discussion Paper 7316, (1973).

[80] N. Veraverbeke, Asymptotic behavior of Wiener-Hopf factors of a random walk, Stoch. proc. Appl. 5, (1977) 27-37.

[81] E. Willekens and J. Teugels, Asymptotic expansions for waiting time probabilities in an M/G/1 queue with long-tailed service times, Queueing Systems 10, (1992) 295-312.

[82] S. Zachary, A note on Veraverbeke's theorem, Queueing Systems 46, (2004) 9-14